

A DNA Metabarcoding Study of a Polyphagous Beetle Dietary Diversity: the Utility of Barcodes and Sequencing Techniques*

Łukasz KAJTOCH

Accepted May 15, 2014

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

Recently developed techniques of DNA barcoding and next-generation sequencing (NGS) overcome previous limitations of evaluation of animal diet composition and together are promising method in molecular ecology. The objective was to compare standard ABI Sanger sequencing with new high throughput sequencing (Illumina MiSeq) technique and the two selected plant barcodes (rbcL gene and trnL intron) in terms of the identification of host plant composition for the selected beetle species – the *Centricnemus leucogrammus* weevil. A comparison of two sequencing techniques showed that NGS (in this case Illumina) gave more exhaustive results than the Sanger method. Moreover, it was proven that a two-locus barcoding systems (rbcL and trnL) is sufficient for host plant identification from DNA isolated from insect bodies, at least at the genus level. A comparison of host plant composition among distant populations revealed that the studied species did not feed uniformly across its range. This probably reveals an ecological adaptation of geographically and genetically isolated populations. These findings, beside broadening basic knowledge on the use of barcoding and sequencing techniques for host plant identifications in insect populations, can have implications for conservation studies and strategies for rare and endangered species.

Key words: Molecular ecology; chloroplast DNA; rbcL gene; trnL intron; phytophagy; *Centricnemus leucogrammus*.

Łukasz KAJTOCH, Institute of Systematics and Evolution of Animals, Polish Academy of Science, Sławkowska 17, 31-049 Kraków, Poland.
E-mail: kajtoch@isez.pan.krakow.pl

The feeding preferences of invertebrates represent a difficult field of research. Most studies have not afforded detailed knowledge about the host plants and feeding habitats of herbivorous arthropods which would lead to an understanding of their complex evolutionary associations and ecological interactions. Moreover, detailed information on the diet of rare and endangered species might be of special value for designing appropriate conservation strategies and management plans (e.g., MARRERO *et al.* 2004). There are several known methods of evaluation of animal diet composition (for review see, e.g., MATHESON *et al.* 2008 and VALENTINI *et al.* 2009a). These methods use either direct observation of foraging insects in nature (e.g., SANDHOLM & PRICE 1962) or in laboratory conditions (e.g., DIECKMANN 1980), or analysis of gut contents and faeces (e.g., HOLECHEK *et al.* 1982; JOHNSON & NICOLSON 2001). Most methods of gut contents and faeces analysis have serious drawbacks related to the quality of plant

material present in samples (fragmentation, digestion) and resolution (possibility to identify plant species). Only the DNA method has been proved to be sufficient for precise identification of host plants and also for quantitative analysis. This method takes advantage of the DNA barcoding concept for species identification (HEBERT *et al.* 2003; MARYAŃSKA-NADACHOWSKA *et al.* 2010; KUBISZ *et al.* 2012a), employed also in ecological and biodiversity studies (VALENTINI *et al.* 2009b; TABERLET *et al.* 2012). Several DNA barcodes have been proposed for land plants, either individually or in combinations (CHASE *et al.* 2007; KRESS & ERICKSON 2007; FAZEKAS *et al.* 2008; HOLLINGSWORTH 2009), and a two-locus barcode has been proposed: the rbcL gene together with the matK gene (CBOL 2009). However, these barcodes may not always be used for the identification of plant species from gut contents as primers for these markers rarely cover a wide spectrum of plant taxonomic units. On the other hand, one

*Supported by grant UMO-2011/01/B/NZ8/01491 from the Polish National Science Centre.

chloroplast intron, trnL (TABERLET *et al.* 1991, 2007; VALENTINI *et al.* 2009a), has been successfully used for host plant barcoding in insects, particularly beetles (JURADO-RIVERA *et al.* 2009; NAVARRO *et al.* 2010; KUBISZ *et al.* 2012b; KAJTOCH *et al.* 2013; KITSON *et al.* 2013). The above studies mostly focused on comparing the sequences gathered from the studied species with the data available from genetic databases such as GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>). This step can also limit the success of identification, as not all plant barcodes have a wide coverage of plant sequences in databases; e.g., the rbcL gene has 126,485 hits, the matK gene 82,631 hits, the trnL intron 160,932 hits and trnH-psbA spacer only 17,287 hits (GenBank, state as of July 2013). These numbers show that the best choices are the trnL intron, the rbcL and the matK genes.

Studies on DNA barcoding of host plants were first based on ABI (Applied Biosystems Inc.) Sanger sequencing; however, this method has serious limitations as it cannot simultaneously generate data for many samples and an additional cloning step is needed for the sequencing of multiple amplicons. This has been recently overcome by next-generation sequencing (NGS) techniques for targeted sequencing studies (HARISMENDY *et al.* 2009; EKBLOM & GALINDO 2011; SHOKRALLA *et al.* 2012). Different NGS techniques have been used for massive sequencing of environmental samples, e.g., for host plant studies (TABERLET *et al.* 2007; SOINONEN *et al.* 2009; POMPANON *et al.* 2012).

Most studies using DNA plant barcodes for host plant identification have been performed on individual species and populations, or even on single individuals (JURADO-RIVERA *et al.* 2009; NAVARRO *et al.* 2010; JUNILLA *et al.* 2010; GARCIA-ROBLEDÓ *et al.* 2013) and on mono- or oligophagous species (STAUDACHER *et al.* 2011; KUBISZ *et al.* 2012b; KAJTOCH *et al.* 2013). So far only KITSON *et al.* (2013) studied polyphagous beetle diet on population level. There are no published studies dealing with host plant identification in a large number of invertebrate specimens from different areas, while only large-scale sampling from many individuals and from different parts of a species range can lead to a comprehensive understanding of the species feeding preferences and behavior. Such an approach is however time- and cost-consuming. It is not always necessary to study many specimens separately to obtain data about the host plant composition of particular populations. It is especially problematic for polyphagous and abundant invertebrates. In such cases, the solution is to analyze their diet at the level of populations instead of individuals. The NGS technique is particularly efficient in this kind of analysis, as it allows for gathering data from many individuals simultaneously.

The xerothermic habitats of central and eastern Europe, which sustain a very high biodiversity of plants (DZWONKO & LOSTER 2007) and insects (LIANA 1987; MAZUR 2001) were selected as the target of this study. Simultaneously, these extrazonal analogs of Eurasian steppes are seriously threatened as a result of their limited and fragmented distribution, as well as anthropogenic transformation and degradation (MICHALIK & ZARZYCKI 1995; MAZUR & KUBISZ 2000; CREMENE *et al.* 2005). In xerothermic habitats, there exist rich assemblages of herbivorous beetles, and especially weevils (Curculionidae) (MAZUR 2001) and leaf-beetles (Chrysomelidae) (WAŚOWSKA *et al.* 2006). Over the past years, great emphasis has been placed on understanding the phylogeography and conservation genetics of selected xerothermic beetle species (KAJTOCH 2011; KAJTOCH *et al.* 2011, 2014; KUBISZ *et al.* 2012b). The species that has been best investigated is the weevil *Centricnemus leucogrammus* (GERMAR 1824), whose genetics has been studied on the basis of several markers (both mtDNA and nucDNA) (LACHOWSKA *et al.* 2006; KAJTOCH *et al.* 2009; KAJTOCH 2011), as well as microsatellites (KAJTOCH *et al.* 2014). These data have shown that *C. leucogrammus* populations can be clustered into several geographically separated clades. It would be interesting to verify if these distinct geographic and genetic units are also distinct ecologically. The species is known to be polyphagous (DIECKMANN 1980; MAZUR 2001); however, its host plants have been identified only via direct observations. These observational data show that it is leave-feeder as imago, however nothing is known about diet of its larva, which probably feed in the root of herbaceous plants as most Entiminae do. The comprehensive knowledge of the zoogeography and genetics of this particular species makes it an excellent subject for the present study.

In this study, the DNA-based approach was used for host plant identification by combining the plant barcoding concept with DNA sequencing techniques. The objective was to compare standard Sanger sequencing with a new high throughput sequencing technique (Illumina MiSeq) in terms of the identification of host plant composition for the selected beetle species – the *C. leucogrammus* weevil. Moreover, both techniques were used for the estimation of differences in host plant composition among four regional groups of this beetle. The third aim of this study was to compare the efficiency of host plant identification between the two selected plant barcodes (the rbcL gene and the trnL intron). The collected data are discussed in respect to the phylogeography of the studied beetle to verify if its distinct geographic and genetic units also have distinct ecological adaptations.

Material and Methods

Sampling

C. leucogrammus was collected in four regions of its distribution in central and eastern Europe: central-southern (CS) Poland (the Małopolska Upland), northern (N) Poland (the Lower Vistula and Lower Oder valleys), western (W) Ukraine (the Podolian Upland) and southern Moravia with south-western Slovakia. The locations of the sampling sites were presented in KAJTOCH *et al.* (2009) and KAJTOCH (2011). In total, 25 specimens were randomly collected from each of these regions in 2010–2013. All specimens were immediately preserved in 99% ethanol and then stored at -20°C . All specimens were identified as *C. leucogrammus* by a specialist entomologist.

Laboratory procedure

Whole beetle bodies were used for DNA isolation (JURADO-RIVERA *et al.* 2009) instead of the time-consuming gut contents preparations proposed by MATHESON *et al.* (2008). Specimens were washed in distilled water before isolation and crushed in test tubes. The insect bodies were subjected to DNA extraction using a NucleoSpin Tissue kit (Macherey-Nagel, Düren, Germany). DNA isolates were used for amplification of three chloroplast barcodes, that is, the genes *matK* and *rbcL* and the intron *trnL*, using the following primers: *matK*472F and *matK*1248R (YU *et al.* 2011), *rbcL*-F1 and *rbcL*-724R (FAY *et al.* 1997), and A49325 and B49863 (TABERLET *et al.* 1991). As *rbcL* and *trnL* gave PCR products from all isolates and *matK* could be amplified in approximately only 70% of the specimens, this last barcode was omitted from further work. PCR conditions and the concentration of reagents were as in the above-mentioned papers. Primers for long fragments (400 bp) of chloroplast DNA (cpDNA) were selected, as the objective of this study was to establish host plants at least at the genus level. Shorter cpDNA barcodes used for host plant identification from degraded samples (e.g., faeces) could also be used, but their resolution is too low and they mainly allow for identification of plants at the family or order level. Moreover, previous works on beetles (JURADO-RIVERA *et al.* 2009; NAVARRO *et al.* 2010; KUBISZ *et al.* 2012b; KAJTOCH *et al.* 2013; GARCIA-ROBLEDO *et al.* 2013; KITSON *et al.* 2013; WALLINGER *et al.* 2013) proved that high quality DNA can be isolated from the guts of insects immediately preserved in ethanol.

Sanger sequencing

Each isolate was amplified separately for *rbcL* and *trnL* barcodes using a Qiagen PCR Core Kit

(Hilden, Germany) and afterwards PCR products were separated on 2% agarose gel. After purification, which was conducted using NucleoSpin Extract II (Macherey-Nagel, Düren, Germany), the PCR products were sequenced using a BigDye Terminator v.3.1. Cycle Sequencing Kit (Applied Biosystems, Carlsbad, CA, USA) and ran on an ABI 3100 Automated Capillary DNA Sequencer.

Illumina sequencing

The same volume (1.0 μL) of each isolate of all 25 specimens from each of the four regions separately was mixed and used as a template for amplification using Phusion® Flash High-Fidelity PCR Master Mix (Waltham, USA), separately for *rbcL* and *trnL* barcodes. Next, the PCR products of *rbcL* and *trnL* barcodes from the same region were mixed. Four samples (one from each region, each containing *trnL* and *rbcL* amplicons) were sent to Genomed company (Warsaw, Poland) for library preparation (with TruSeq DNA LT Sample Preparation Kit (FC-121-2001, Illumina), sample tagging, and sequencing in a 1/8 run of Illumina's compact MiSeq system with 2 \times 250 bp read lengths (using a MiSeq Reagent Kit v2). I decided not to tag individuals as this would be too costly, while the aim of this study was analysis at the population level.

Data analysis

Sanger

Sequences were checked by eye using BioEdit v.7.0.5.2 (HALL 1999). Only sequences of good quality fragments, longer than 400 bp (*trnL*) or 650 bp (*rbcL*), were used for further analysis.

Sequences were stored in FASTA files and used for identification of host plants with MEGABLAST (Basic Local Alignment Search Tool, ALTSCHUL *et al.* 1990) searching a nucleotide database using nucleotide queries (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Only sequences of a query coverage larger than 95%, E value = 0, and a maximum identity larger than 98% were retrieved. For each sequence, the most similar hit was noted (species name and GenBank accession number – not presented).

Illumina

Illumina paired reads were checked for their quality with use of cutadapt (Martin 2011) and low quality reads (–q) were trimmed before further analysis. Subsequently, automatically joined forward and reverse paired reads were overlapped (when at least 50 bp of homologous sequences were found between forward and reverse reads)

with use of CLCGenomicWorkbench package (<http://www.clcbio.com/products/clc-genomics-workbench/>). Next, the prepared FASTA files were compared to rbcL and trnL sequence databases downloaded from the GenBank. Only sequences which apparently belonged to one of these two groups were selected (with use of blast2seq option in BLAST) and FASTA files were prepared separately for rbcL and trnL barcodes (8 files in total). Next, sequences were filtered according to their length using GALAXY (GOECKS *et al.* 2010) and only sequences longer than 350 bp were retrieved for further analysis. A 350 bp sequence length threshold was used, as below this threshold most MEGABLAST hits showed E values much larger than 0 for both studied DNA fragments. Subsequently, two searching strategies were implemented.

First, sequences in each file were collapsed using GALAXY and a new database (FASTA file) was built on unique sequence variants. These files were used for host plant identification using MEGABLAST in two ways. In the first method, all files were compared with the previously obtained FASTA files of Sanger sequences (separately for rbcL and trnL) using the “align two or more sequences” option in MEGABLAST. In the second method – Illumina files were compared against sequences deposited in NCBI GenBank, also using MEGABLAST. Similarly as for Sanger files, only sequences of query coverage larger than 95%, and a maximum identity larger than 98% were collected; however, not all of these sequences had E value = 0 due to their shorter length than the Sanger reads.

The second option was to use MEGAN V4.70.4 (HUSON *et al.* 2011) for comparing DNA sequences against databases of known sequences using BLASTn and then for computing and exploring the taxonomical content of the data set employing NCBI (National Center for Biotechnology Information, <http://www.ncbi.nlm.nih.gov/>) taxonomy to summarize and order the results. The program uses a simple algorithm that assigns each read to the lowest common ancestor (LCA) of the set of taxa that it identified in the comparison. As a result, species-specific sequences are assigned to taxa near the leaves of the NCBI tree, whereas widely conserved sequences are assigned to high-order taxa closer to the root.

The main goal of these comparisons was to establish a list of host plant genera, but not in all cases (both for Sanger and Illumina sequences) was it possible to identify individual genera; then two most probable (and closely related) plant genera were used instead as host plant taxa. For Sanger and Illumina data, information was collected about the host plant species composition within each of

studied regions as well as for all the studied populations. Moreover, for Sanger data, an estimate of frequencies of host plant genera (number of hits for 25 sequenced beetles) was calculated separately for each region. This analysis was done only for rough estimation of host plant frequencies taking into consideration the limitations of Sanger sequencing. Data from Illumina were not used for quantitative analyses as too many variables could influence the proportion of reads of particular host plant taxa (e.g., DNA concentration in mixed isolates, not even the rate of amplification of particular taxa due to priming mismatches, Illumina sequencing, and bioinformatic steps).

Data about host plant species composition among four regions, between two barcodes and between two sequencing methods were compared using statistics (t-tests, χ^2 tests, Analysis of Variance, Principal Component Analysis) in Statistica 11 package (StatSoft).

The lists of host plants (obtained from both markers and both sequencing techniques) were compared against a database of xerothermic plant species of central and eastern Europe that was prepared in a concurrent project (Kajtoch Ł., Heise W.) and with a full list of plant species of Poland (<http://www.atlas-roslin.pl/>). This step was done to determine the most probable host plant species for the *C. leucogrammus* weevil and verify how many of the identified host plants were indeed xerothermic.

In addition, the collected data are discussed in respect to the phylogeography and population genetics of the studied beetle (KAJTOCH *et al.* 2009, 2014; KAJTOCH 2011).

Results

PCR success

All 100 specimens (25 from each of the four regions, collected during the vegetation season and under good weather conditions) gave PCR products using both trnL and rbcL barcodes.

RbcL barcodes implemented under Sanger sequencing resulted in 680-700 bp long amplicons, while trnL amplicons were much more variable (400-600 bp long) (Fig. 1).

In total Illumina sequencing resulted in 104654 paired reads of 2x250 bp length each. There were 24485 Illumina rbcL reads of joined forward and reverse sequences, among which 5222 reads could be overlapped and 624 reads were longer than 350 bp. TrnL Illumina sequencing gave many more reads: 80169 joined, 11426 overlapped, and 2903 longer than 350 bp.

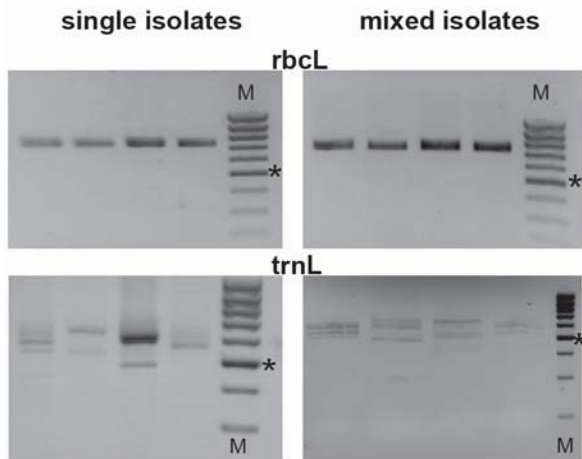


Fig. 1. Gel electrophoresis of *rbcL* and *trnL* PCR products for chosen single isolates of *Centricnemus leucogrammus* weevil and mixed isolates from four studied regions of species distribution (from the left: central-south Poland, north Poland, west Ukraine, Slovakia-Moravia). * – 500 bp marker (M) band.

General data

Twenty plant genera (in some cases pairs of closely related genera) were found using Sanger sequencing and 30 genera using Illumina sequencing. In particular regions, we found 12, 9, 11, and 6 genera (Sanger) and 19, 17, 19, and 20 genera (Illumina) for CS Poland, N Poland, W Ukraine, and Moravia-Slovakia, respectively. Using data from both Sanger and Illumina sequencing, 20, 17, 19, and 20 plant genera were found in these regions, respectively. In total, 30 plant genera were found.

Considering plant barcodes, 29 genera were identified using the *rbcL* gene (19 by Sanger or 29 by Illumina), and 27 using the *trnL* intron (20 and 24, respectively). In respect of the regions, the figures for *rbcL* were as follows: 11 identified by Sanger and 17 by Illumina in CS Poland; 8 and 16 in N Poland; 11 and 16 in W Ukraine, and 6 and 16 in Moravia-Slovakia, respectively. The corresponding figures for *trnL* were 12 and 16 genera in CS Poland, 9 and 13 in N Poland, 11 and 15 in W Ukraine, and 6 and 15 in Moravia-Slovakia, respectively.

Sanger

Considering host plant barcoding, using only the Sanger method, we found that for 15-25% (*rbcL*) and 15-20% (*trnL*) of specimens host plants could not be determined due to mixed sequences obtained in ABI chromatograms (Table 1, Fig. 1). These specimens apparently fed on more than one host plant. Most individuals gave single bands of PCR products on agarose gel; however, some

(those described above as “mixed”) gave two or three bands or smears, which could not be separated in most cases. No of such “mixed” individuals could be resolved in *rbcL*; however, in up to 12% of such specimens (0-3 per region) I was able to clearly separate PCR bands of *trnL* sequences, which were then cleaned and successfully sequenced (Fig. 1).

There was no difference in the total effect of amplification and sequencing success between *rbcL* and *trnL* barcodes ($t_{99}=0.00$, $P=1.00$), or when the amplification and sequencing success for the two markers was considered in individual regions (all $P>0.05$). On the other hand, when considering host plant composition and frequency of particular genera in the four regions, we found significant differences between all pairs of these regions at the level of family (all $P>0.05$) and genus (all $P>0.05$, only between CE and N Poland $P=0.049$) (details are given in Table 2).

Illumina and Sanger/Illumina comparison

Generally, Illumina sequencing revealed approximately 30% more host plant genera than Sanger – Tables 1, 2, and Fig. 2). Differences in host plant detection between *rbcL* and *trnL* barcodes were significantly higher when using Illumina ($t_{20}=2.25$, $P=0.032$).

Taxon assignment using directly MEGABLAST with the identification of the most similar sequences from GenBank and using the LCA algorithm implemented in MEGAN gave similar results, i.e., the same plant genera (or higher taxa) were identified using both methods for particular regions using either *rbcL* or *trnL* barcodes.

All the plant genera identified using Sanger were also detected by Illumina sequencing, but not all by both barcodes. The exceptions were *Trifolium* and *Medicago* (absent from *rbcL*) and *Thalictrum* and *Sanguisorba* (absent from *trnL*). There were also some differences in the distribution of plant genera in the studied regions; typically some genera found to be present in all or almost all regions in accordance with Illumina were detected in only some of the regions in accordance with Sanger (e.g., *Thymus*, *Inula*, *Genista* and *Dorycnium/Lotus* identified by *rbcL* and *Inula*, *Dorycnium/Lotus*, and *Medicago* identified by *trnL*) (Figs 2 and 3). Moreover, Illumina sequencing led to the identification of 10 genera more than Sanger: *Potentilla/Fragaria*, *Crataegus*, *Cytisus*, *Triticum/Arrhenatherum*, *Brachypodium*, *Allium*, *Corylus/Betula*, *Rumex*, *Helianthemum*, *Datura/Lycopersicon* (10 by *rbcL* and 6 by *trnL*) (Tables 1 and Fig. 2).

There were significant differences in total host plant composition between Sanger and Illumina in

Table 1

Composition of *Centricnemus leucogrammus* host plants in respect to DNA barcode type (rbcL and trnL) and sequencing technique (Sanger and Illumina) in four regions of species distribution: S – Sanger sequencing, I – Illumina sequencing, CSP – central-south Poland, NP – north Poland, WU – western Ukraine, SM – Slovakia-Moravia

plant taxa	rbcL						trnL						Sanger					Illumina					Sanger+Illumina																				
	CSP		NP		WU		SM		all regions		NP		WU		SM		all regions		rbcL + trnL		rbcL		NP		WU		SM		all regions		rbcL + trnL		rbcL		NP		WU		SM		all regions		
	S	I	S	I	S	I	S	I	S	I	S	I	S	I	S	I	S	I	S	I	CSP	NP	WU	SM	all regions	CSP	NP	WU	SM	all regions	CSP	NP	WU	SM	all regions	CSP	NP	WU	SM	all regions			
<i>Daucus/Heracleum</i>	?	+	+	+	+	+	+	+	+	+										+	+			1	+	+	+	+	1	+	+	+	+	+	+	+	1	+	+	+	+	1	
<i>Imula</i>	+	+	+	+						+										+	+			1	+	+	+	+	4	+	+	+	+	+	+	+	+	+	4	+	+	+	4
<i>Hieracium/Crepis</i>	+	+	+	+						+										+	+			3	+	+	+	+	4	+	+	+	+	+	+	+	+	+	4	+	+	+	4
<i>Achillea</i>	+	+	+	+						+										+	+			3	+	+	+	+	4	+	+	+	+	+	+	+	+	+	4	+	+	+	4
<i>Centaurea</i>	+	+	+	+						+										+	+			3	+	+	+	+	4	+	+	+	+	+	+	+	+	+	4	+	+	+	4
<i>Salvia</i>	+	+	+	+						+										+	+			4	+	+	+	+	4	+	+	+	+	+	+	+	+	+	4	+	+	+	4
<i>Thymus</i>	+	+	+	+						+										+	+			2	+	+	+	+	4	+	+	+	+	+	+	+	+	+	4	+	+	+	4
<i>Datura/Lycopersicon</i>	+																			+	+			0	+	+	+	+	2	+	+	+	+	+	+	+	+	2	+	+	+	2	
<i>Asperula</i>	+									+										+	+			1	+	+	+	+	1	+	+	+	+	+	+	+	+	1	+	+	+	1	
<i>Rumex</i>																				+	+			0	+	+	+	+	2	+	+	+	+	+	+	+	+	2	+	+	+	2	
<i>Helianthemum</i>										+										+	+			0	+	+	+	+	1	+	+	+	+	+	+	+	+	1	+	+	+	1	
<i>Prunus</i>	+	+	+	+						+										+	+			3	+	+	+	+	4	+	+	+	+	+	+	+	+	4	+	+	+	4	
<i>Sanguisorba</i>	+									+										+	+			2	+	+	+	+	1	+	+	+	+	+	+	+	+	2	+	+	+	2	
<i>Filipendula</i>		+	+	+						+										+	+			3	+	+	+	+	4	+	+	+	+	+	+	+	+	4	+	+	+	4	
<i>Rosa</i>			+	+						+										+	+			1	+	+	+	+	1	+	+	+	+	+	+	+	1	+	+	+	1		
<i>Rubus</i>		+																		+	+			0	+	+	+	+	1	+	+	+	+	+	+	+	1	+	+	+	1		
<i>Potentilla/Fragaria</i>																				+	+			0	+	+	+	+	2	+	+	+	+	+	+	+	2	+	+	+	2		
<i>Crataegus</i>																				+	+			0	+	+	+	+	1	+	+	+	+	+	+	+	1	+	+	+	1		
<i>Cytisus</i>	+	+	+	+						+										+	+			2	+	+	+	+	3	+	+	+	+	+	+	+	3	+	+	+	3		
<i>Genista</i>	+	+	+	+						+										+	+			1	+	+	+	+	3	+	+	+	+	+	+	+	3	+	+	+	3		
<i>Onobrychis/Hedysarum</i>	+	+	+	+						+										+	+			2	+	+	+	+	4	+	+	+	+	+	+	+	4	+	+	+	4		
<i>Dorycnium/Lotus</i>		+	+	+						+										+	+			1	+	+	+	+	2	+	+	+	+	+	+	+	2	+	+	+	2		
<i>Trifolium</i>		+	+	+						+										+	+			1	+	+	+	+	4	+	+	+	+	+	+	+	4	+	+	+	4		
<i>Medicago</i>																				+	+			2	+	+	+	+	2	+	+	+	+	+	+	+	2	+	+	+	2		
<i>Corylus/Betula</i>	+																			+	+			0	+	+	+	+	4	+	+	+	+	+	+	+	4	+	+	+	4		
<i>Thalictrum</i>																				+	+			1	+	+	+	+	2	+	+	+	+	+	+	+	2	+	+	+	2		
<i>Anemone</i>																				+	+			1	+	+	+	+	1	+	+	+	+	+	+	+	1	+	+	+	1		
<i>Triticum/Arrhenatherum</i>	+									+										+	+			0	+	+	+	+	1	+	+	+	+	+	+	+	1	+	+	+	1		
<i>Brachypodium</i>																				+	+			0	+	+	+	+	1	+	+	+	+	+	+	+	1	+	+	+	1		
<i>Allium</i>	+																			+	+			0	+	+	+	+	4	+	+	+	+	+	+	+	4	+	+	+	4		
Σ	11	17	8	16	11	16	6	16	19	29	12	16	9	13	11	15	6	15	20	24	12	9	11	6	19	17	19	20	20	17	19	20	19	17	19	20	20	17	19	20	20		

Table 2

Statistics for comparison host plant frequencies, identified based on Sanger sequencing, among four regions of *Centricnemus leucogrammus* distribution (χ^2_{24} below diagonals, P values above diagonals) and Spearman's coefficient of rank correlation (r_s) of host plant composition among regional populations of *Centricnemus leucogrammus* based on concentrated data (Sanger and Illumina), P values above diagonal

Region	CS Poland	N Poland	W Ukraine	Slovakia-Moravia
χ^2 - family level				
CS Poland	-	0.030	0.008	0.014
N Poland	13.94	-	0.001	0.018
W Ukraine	17.52	17.83	-	0.000
Slovakia-Moravia	14.36	11.92	27.00	-
χ^2 - genus level				
CS Poland	-	0.017	0.049	0.015
N Poland	30.31	-	0.011	0.034
W Ukraine	28.99	28.85	-	0.001
Slovakia-Moravia	27.87	19.57	33.99	-
Spearman's coefficient of rank correlation				
CS Poland	1.00	0.095	0.196	0.250
N Poland	0.10	1.00	0.312	0.238
W Ukraine	0.20	0.31	1.00	0.342
Slovakia-Moravia	0.25	0.24	0.34	1.00

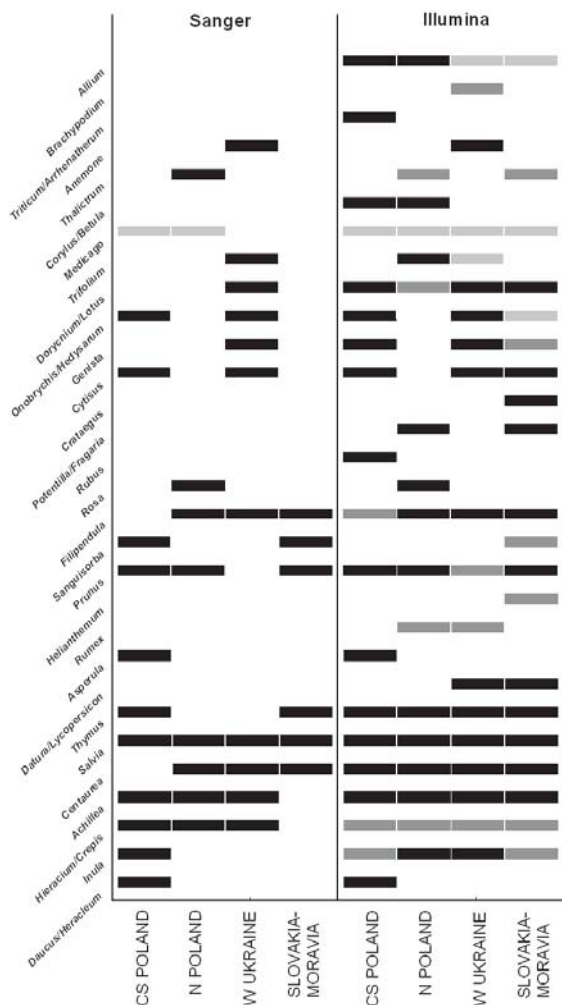


Fig. 2. *Centricnemus leucogrammus* host plant composition revealed by Sanger and Illumina sequencings of plant barcodes and differences between these two methods in four regional populations of studied weevil. Black rectangles – rbcL+trnL, dark grey rectangles – rbcL, light grey rectangles – trnL.

respect of the rbcL barcode ($t_{29}=-4.28, P=0.0002$), but not in respect of trnL ($t_{29}=-1.64, P=0.1098$).

Host plants

In total, the host plants identified for *C. leucogrammus* belong to 13 orders, 13 families and 19 subfamilies. Most orders and families were represented by single genera. The greatest number of genera were found in Rosaceae, Fabaceae, and Asteraceae, while Lamiaceae, Ranunculaceae, and Poaceae were represented by two genera each. At

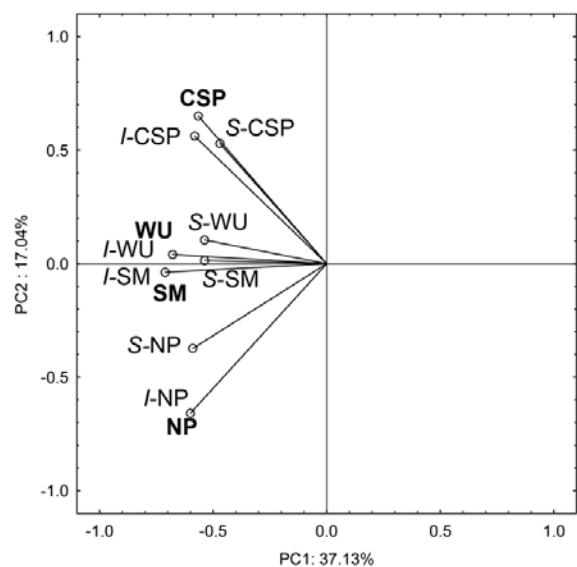


Fig. 3. Plot of the first two principal components (PC1 and PC2) of an analysis of *Centricnemus leucogrammus* host plant composition revealed by concentrated data of Sanger (S) and Illumina (I) sequencings for four regional populations of studied weevil: CSP – central-south Poland, NP – north Poland, WU – west Ukraine, SM – Slovakia-Moravia.

Table 3

Systematics of host plants of *Centricnemus leucogrammus* identified base on Sanger and Illumina sequencing of two plant barcodes (rbcL and trnL) and most probable representatives of central-east European flora. In bold - species characteristic for xerothermic turfs and steppes, N - host plants identified by previous observation in nature (Mazur 2001), F - plants eaten by *C. leucogrammus* during feeding experiments (Dieckmann 1980)

Order	Family	Subfamily	Tribe	Plant species	N	F
Apiales	Apiaceae	Apiioideae		<i>Daucus carota</i> , <i>Heracleum sphondylium</i>		
Asterales	Asteraceae	Asteroideae	Inuleae	<i>Inula ensifolia</i> , <i>I. hirta</i>	+	
Asterales	Asteraceae	Cichorioideae	Cichorieae	many <i>Hieracium</i> species, <i>Crepis praemorsa</i>	+	+
Asterales	Asteraceae	Asteroideae	Anthemideae	<i>Achillea collina</i> , <i>A. pannonica</i> , <i>A. setacea</i> , <i>A. millefolium</i>		+
Asterales	Asteraceae	Carduoideae	Cardueae	<i>Centaurea scabiosa</i> , <i>C. stoebe</i>		
Lamiales	Lamiaceae	Salvioideae	Salvieae	<i>Salvia nemorosa</i> , <i>S. pratensis</i> , <i>S. verticillata</i>	+	+
Lamiales	Lamiaceae		Thymaeae	many <i>Thymus</i> species		
Solanales	Solanaceae	Daturoideae	Datureae	<i>Datura stramonium</i>		
Gentianales	Rubiaceae	Rubieae	Asperuleae	<i>Asperula cynanchica</i> , <i>A. tinctoria</i>		
Caryophyllales	Polygonaceae	Polygonoideae	Rumiceae	<i>Rumex ucranicus</i> , <i>R. acetosella</i>		
Malvales	Cistaceae			<i>Helianthemum nummularium</i>		+
Rosales	Rosaceae	Spiraeoideae	Amygdaleae	<i>Prunus spinosa</i>		
Rosales	Rosaceae	Rosoideae	Sanguisorbeae	<i>Sanguisorba minor</i>		
Rosales	Rosaceae	Rosoideae	Ulmarieae	<i>Filipendula vulgaris</i>		
Rosales	Rosaceae	Rosoideae		<i>Rosa canina</i>		
Rosales	Rosaceae	Rosoideae	Rubeae	many <i>Rubus</i> species		
Rosales	Rosaceae	Rosoideae	Potentilleae	<i>Potentilla arenaria</i> , <i>P. argentea</i> , <i>Fragaria viridis</i>	+	+
Rosales	Rosaceae	Spiraeoideae	Pyreae	many <i>Crataegus</i> species		+
Fabales	Fabaceae	Faboideae	Genisteae	<i>Cytisus scoparius</i>		
Fabales	Fabaceae	Faboideae	Genisteae	<i>Genista tinctoria</i>		
Fabales	Fabaceae	Faboideae	Hedysareae	<i>Onobrychis viciifolia</i>		
Fabales	Fabaceae	Faboideae	Loteae	<i>Dorycnium pentaphyllum</i> , <i>Lotus corniculatus</i>		
Fabales	Fabaceae	Faboideae	Trifolieae	<i>Trifolium pratense</i> , <i>Trifolium pannonicum</i>		
Fabales	Fabaceae	Faboideae	Trifolieae	<i>Medicago falcata</i> , <i>M. lupulina</i>		+
Fagales	Betulaceae	Coryloideae/Betuloideae	Coryleae/Betuleae	<i>Corylus avellana</i> , many <i>Betula</i> species		
Ranunculales	Ranunculaceae	Thalictroideae		<i>Thalictrum minus</i>		
Ranunculales	Ranunculaceae		Anemoneae	<i>Anemone sylvestris</i>		+
Poales	Poaceae	Pooideae	Triticeae/Poeae	<i>Arrhenatherum elatius</i> , many <i>Triticum</i> species		
Poales	Poaceae	Pooideae	Brachypodieae	<i>Brachypodium pinnatum</i>		
Asparagales	Amaryllidaceae	Allioideae		<i>Allium montanum</i>		

the subfamily level, Rosoideae and Faboideae were most frequently found. In terms of tribes, more than one genus was found for Genisteae, Trifolieae, and also for Cichorieae, Potentilleae, and Loteae (among which genera were not distinguishable with the present data).

A full list of *C. leucogrammus* host plants is presented in Table 3. In particular regions, 17-20 plant genera were found; however, their composition varied. Eleven plant genera were present in all *C. leucogrammus*-inhabited regions (*Inula*, *Hieracium*/*Crepis*, *Achillea*, *Centaurea*, *Salvia*, *Thymus*, *Prunus*, *Filipendula*, *Dorycnium*/*Lotus*, *Medicago*, *Allium*), whereas 5 were found only in single regions (*Daucus*/*Heracleum*, *Asperula*, *Rubus*, and *Triticum*/*Arrhenatherum* in CS Poland; *Rosa* in N Poland; *Brachypodium* and *Anemone* in W Ukraine; and *Helianthemum* and *Crataegus* in Moravia-Slovakia). The other plant genera were present in 2 or 3 regions.

Differences in host plant composition were significant among regions both for the rbcL barcode (AMOVA=10.629, P=0.014) and the trnL barcode (AMOVA=8.472, P=0.037), as well as for rbcL + trnL (AMOVA=9.296, P=0.025). Principal component analysis (PCA) conducted on the Sanger and Illumina data concerning host plant composition for the four regions of *C. leucogrammus* distribution showed that both methods gave generally concordant results within regions (Fig. 3). On the other hand, PCA revealed that populations from each of studied region had a distinct host plant composition, the most similar being populations from W Ukraine and Slovakia-Moravia (Fig. 3). This was additionally supported by a lack of significant correlations between host plant compositions (concentrated data) among all the studied regional populations (all P of Spearman's coefficient of rank correlation >> 0.05, Table 2).

Among the 30 plant taxa identified as host plants for *C. leucogrammus*, 25 had xerothermic representatives in central and eastern Europe (Table 3). Other plants are either frequent on xerothermic turfs and in similar habitats (e.g., sandy turfs), ruderal (some also alien and anthropophilic) or widespread and eurytopic (Table 3).

Discussion

The presented study is the first to precisely characterize the diet of polyphagous taxa from different parts of its range, and also to compare the utility of two sequencing techniques (Sanger and NGS) and two plant barcodes (rbcL gene and trnL intron) for host plant barcoding.

DNA barcoding with two sequencing techniques showed, as it could be expected, that NGS sequencing (in this case Illumina MiSeq) works much better than traditional Sanger sequencing. This last method failed to identify about 10 plant genera (about 30%) present in the diet of *C. leucogrammus* and, moreover, in several cases, it failed to detect some plant genera in particular populations. This was caused by the presence of DNA templates from multiple host plants in some individuals. These individuals could not be sequenced using Sanger without a time-consuming and costly cloning step (not used in this study) for the rbcL gene (as this DNA barcode has almost the same length in most of plants). I was able to investigate only a fraction of individuals with mixed plant DNA after PCR band extraction from agarose gel (as PCR products obtained with trnL from different plant taxa are often of different length). On the other hand, Illumina sequencing led to the identification of all plant genera identified by Sanger, and, additionally, of other plants missed by Sanger. It is worth noting that Illumina sequencing performed on mixed isolates from many individuals does not enable quantitative analysis of host plant frequencies in particular individuals. This can be examined via Sanger sequencing, but due to the limitations of this method and false negative results, quantitative analysis on Sanger data should be treated as an approximation. Illumina sequencing could be used for the sequencing of host plants for individuals using special sample tagging. However, this would be more costly and unnecessary in the case of studies on abundant species, such as most invertebrates. False negative results of Sanger sequencing are easy to explain as resulting from unsuccessful sequencing of mixed isolates. On the other hand, the few examples of Illumina false negative results could be attributed to either problems with DNA amplification on mixed templates (as some plant species could have

mismatches in the priming sites of universal primers), the sequencing step (differences in DNA concentration), or the bioinformatic sorting of data (overlooking of some data, e.g., as a result of a high threshold for sequence length used in this study). It is also interesting to note that the results of the two barcodes used in this study not always led to identical conclusions. First, sometimes one or the other barcode (rbcL or trnL) failed to amplify DNA (probably as a result of the problems mentioned above) or failed to identify the same plant taxa (too low resolution). Second, for some individuals, different host plants were identified using Sanger sequencing, even though they were always very closely taxonomically related. This is probably caused by problems in the assembling step, which regardless of the method, compared data with GenBank sequences. In some cases, sequences deposited in GenBank did not allow for precise recognition of plant genera, as the barcodes fitted more than one plant genus. That is why, in such ecological studies, special barcoding of local vegetation is recommended (VALENTINI *et al.* 2009a; SOINONEN *et al.* 2009; POMPANON *et al.* 2012). Actually, DNA barcoding of central and eastern European xerothermic plants is in progress (Kajtoch Ł., Heise W.).

Comparing two plant barcodes (rbcL and trnL), a higher sequencing success was observed in Illumina, but only for the rbcL gene. The TrnL intron also showed a higher number of plant identifications in Illumina than in Sanger, but these differences were not statistically significant. As was mentioned above for the Sanger method, generally the trnL barcode allowed for the identification of more host plants due to the possibility of sequencing mixed DNA templates after gel electrophoresis. This finding suggests that in studies where NGS cannot be used (because of either technical or economic reasons), it would be better to choose trnL as a barcode. Indeed, reports from other ecological studies also recommend the trnL intron (TABERLET *et al.* 2007; VALENTINI *et al.* 2009a) over rbcL or matK genes (CBOL 2009).

From the biological point of view it is interesting to analyze the collected data in respect of host plant composition, especially in terms of differences between the populations of the studied weevil. This study confirms that *C. leucogrammus* is polyphagous but feeds on many more plant genera than it was expected. Direct observations (either in nature or in laboratory) suggested that this species fed on 14 plant genera (*Inula*, *Potentilla*, *Salvia*, *Adonis*, *Anemone*, *Hieracium*, *Helianthemum*, *Crataegus*, *Medicago*, *Anthyllis*, *Coronilla*, *Origanum*, *Syringa*, *Achillea*) (DIECKMANN 1980; MAZUR 2001). Among these plants, DNA barcoding confirmed 9 genera to be host plants. Others (*Adonis*,

Anthyllis, *Coronilla*, *Origanum*, *Syringa*) belong to plant families frequently recorded as part of the *C. leucogrammus* diet (Fabaceae, Lamiaceae, Ranunculaceae). Furthermore, some of these genera are closely related with those identified by DNA barcoding (*Origanum* with *Thymus*, *Anthyllis* and *Coronilla* with *Lotus/Dorycnium*, *Adonis* with *Anemone*), so they might possibly be misidentified by DNA sequences, or the studied species might feed on both of these closely related genera, which however, could not be determined using plant barcodes. As *C. leucogrammus* is a species exclusively inhabiting xerothermic habitats (MAZUR 2001), it is not surprising that 83% of its host plants are related with this kind of environment. This species is flightless and is always collected in the herbaceous layer of xerothermic turfs. Barcoding confirms that while it feeds almost exclusively on herbaceous plants, it can also feed on some bushes (*Rosa*, *Crataegus*, *Prunus*, *Cytisus*), and even occasionally also on trees (*Corylus/Betula*), perhaps taking advantage fallen leaves. One of the most interesting results of this study is the finding that insects do not need to have the same diet in different areas of their distribution. In particular regions of *C. leucogrammus* range: central-south Poland, northern Poland, western Ukraine, and Slovakia-Moravia, a similar number of host plant genera were found (from 17 to 20, so 55-65% of all identified host plants). Some plant genera were present in all or almost all regions and they probably constitute the most important and constant food source (*Salvia*, *Thymus*, *Centaurea*, *Achillea*, *Inula*, *Hieracium*, *Prunus*, *Filipendula*, *Lotus*, *Medicago*, *Allium*). However, in all of these regions, the host plant composition varied significantly. While this result could certainly be caused by some research errors (sampling, laboratory, or bioinformatic steps), a more probable explanation is that, indeed, in each region this species prefers to feed on different host plants, as all populations were sampled and studied in the same way. Differences in host plant composition among populations can be explained in at least two ways. First, plant distribution varies. However, all of the identified host plant genera have some representatives in all areas of the studied species range. Some exceptions are xerothermic species like e.g. *Rumex ucranicus* (present only in Ukraine and N Poland) and the genus *Dorycnium* (absent in N Poland). The other explanation involves different ecological adaptations of *C. leucogrammus* populations to different host plants, which could be related to the geographic isolation of populations and their genetic distinctiveness (KAJTOCH 2011; KAJTOCH *et al.* 2009, 2012; KAJTOCH & LACHOWSKA-CIERLIK 2009). If so, the conservation units previously identified for this species (KAJTOCH 2011) can be additionally supported by ecological data, and this

may allow for the identification of units under the Adaptive Evolutionary Conservation (AEC) concept (FRASER & BERNATCHEZ 2001), which integrates Evolutionary Units (RYDER 1986) with different criteria, e.g., ecological features.

The presented study is the first example of host plant identification for different populations of a polyphagous insect (beetle). A comparison of two sequencing techniques showed that NGS (in this case Illumina) gave more exhaustive results than the Sanger method. Moreover, it was proven that a two-locus barcoding systems (*rbcL* and *trnL*) is sufficient for host plant identification from DNA isolated from insect bodies, at least at the genus level. A comparison of host plant composition among distant populations revealed that the studied species did not feed uniformly across its range. This probably reveals an ecological adaptation of geographically and genetically isolated populations. These findings, beside broadening basic knowledge on the use of barcoding and sequencing techniques for host plant identifications in insect populations, can have implications for conservation studies and strategies for rare and endangered species. Precise identification of feeding preferences and behavior could be very important for planning conservation and management of populations and habitats. Without detailed data about host plants, it would be impossible to efficiently protect some herbivorous species and whole insect assemblages. This should be especially important for habitats sustaining very rich flora and fauna, such as the xerothermic habitats of central and eastern Europe.

Acknowledgements

I would like to express my sincerest gratitude to Waldemar HEISE for assistance in botanical data interpretation, Daniel KUBISZ and Miłosz A. MAZUR for help in sampling, Wiesław BABIK and Michał STUGLIK for useful advices to bioinformatic analyses, Tomasz FIGARSKI for help in statistics and anonymous reviewers for their helpful comments to previous version of this article.

References

- ALTSCHUL S.F., GISH W., MILLER W., MYERS E.W., LIPMAN D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**: 403-410.
- CBOL PLANT WORKING GROUP 2009. A DNA barcode for land plants. *PNAS* **106**: 12794-12797.
- CHASE M.W., COWAN R.S., HOLLINGSWORTH P.M., VAN DEN BERG C., MADRINAN S., PETERSEN G., SEBERG O., JORGENSEN T., CAMERON K.M., CARINE M., PEDERSEN N., HEDDERSON T.A.J., CONRAD F., SALAZAR G.A., RICHARDSON J.E., HOLLINGSWORTH M.L., BARRACLOUGH T.G., KELLY L.,

- WILKINSON M. 2007. A proposal for a standardised protocol to barcode all land plants. *Taxon* **56**: 295-299.
- CREMENE C., GROZA G., RAKOSY L., SCHILEYKO A., BAUR A., ERHARDT A., BAUR B. 2005. Alterations of steppe-like grasslands in Eastern Europe: a threat to regional biodiversity hotspots. *Conserv. Biol.* **19**: 1606-1618.
- DIECKMANN L. 1980. Beiträge zur Insektenfauna der DDR: Coleoptera – Curculionidae (Brachycerinae, otiorhynchinae, Brachyderinae). *Beiträge zur Entomologie Berlin* **30**: 145-310.
- DZWONKO Z., LOSTER S. 2007. A functional analysis of vegetation dynamics in abandoned and restored limestone grasslands. *J. Veg. Sci.* **18**: 203-212.
- EKBLOM R., GALINDO J. 2011. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* **107**: 1-15.
- FAY M.F., SWENSEN S.M., CHASE M.W. 1997. Taxonomic affinities of *Medusagyne oppositifolia* (Medusagynaceae). *Kew Bull* **52**: 111-120.
- FAZEKAS A.J., BURGESS K.S., KESANAKURTI P.R., PRASAD R., GRAHAM S.W., NEWMASER S.G., HUSBAND B.C., PERCY D.M., HAJIBABAEI M., BARRETT S.C.H. 2008. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLOS ONE* **3**, e2802.
- FRASER D.J., BERNATCHEZ L. 2001. Adaptive evolutionary conservation: towards a unified concept for defining conservation units. *Mol. Eco.* **10**: 2741-2752.
- GARCIA-ROBLEDO C., ERICKSON D.L., STAINES C.L., ERWIN T.L., KRESS W.J. 2013. Tropical plant-herbivore networks: Reconstructing species interactions using DNA barcodes. *PLoS ONE*, **8**: e52967.
- GOECKS J., NEKRUTENKO A., TAYLOR J., The Galaxy Team. 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* **11**: R86.
- HALL T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**: 95-98.
- HARISMENDY O., NG P.C., STRAUSBERG R.L., WANG X., STOCKWELL T.B., BEESON K.Y., SCHORK N.J., MURRAY S.S., TOPOI E.J., LEVY S., FRAZER K.A. 2009. Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol.* **10**, R32PMID 19327155.
- HEBERT P.D.N., RATNASINGHAM S., DEWAARD J.R. 2003. Barcoding animal life: cytochrome oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. B: Biol. Sci.* **270**: 96-99.
- HOLECHEK J.L., VAVRA M., PIEPER R.D. 1982. Botanical composition determination of range diets: a review. *J. Range Manag.* **35**: 309-315.
- HOLLINGSWORTH M.L., CLARK A., FORREST L.L., RICHARDSON J.R., PENNINGTON R.T., PRASAD R., GRAHAM S. W., NEWMASER S.G., HUSBAND B.C., PERCY D.M., HAJIBABAEI M., BARRETT S.C.H. 2009. Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. *Mol. Eco. Res.* **9**: 439-457.
- HUSON D.H., MITRA S., WEBER N., RUSCHEWEYH H., SCHUSTER S.C. 2011. Integrative analysis of environmental sequences using MEGAN4. *Genome Res* **21**: 1552-1560.
- JOHNSON S.A., NICOLSON S.W. 2001. Pollen digestion in flower-feeding Scarabaeidae: protea beetles (Cetoniini) and monkey beetles (Hopliini). *J. Ins. Physiol.* **47**: 725-733.
- JUNNILA A., MÜLLER G.C., SCHLEIN Y. 2010. Species identification of plant tissues from the gut of *An sergentii* by DNA analysis. *Acta Trop.* **115**: 227-233.
- JURADO-RIVERA J.A., VOGLER A.P., REID C.A.M., PETITPIERRE E., GÓMEZ-ZURITA J. 2009. DNA barcoding insect-hostplant associations. *Proc. R. Soc. B: Biol. Sci.* **276**: 639-648.
- KAJTOCH Ł., LACHOWSKA-CIERLIK D., MAZUR M. 2009. Genetic diversity of xerothermic weevils *Polydrusus inustus* and *Centricnemus leucogrammus* (Coleoptera: Curculionidae) in central Europe. *Eur. J. Entomol.* **106**: 325-334.
- KAJTOCH Ł., LACHOWSKA-CIERLIK D. 2009. Genetic constitution of parthenogenetic form of *Polydrusus inustus* (Coleoptera: Curculionidae) – hints of hybrid origin and recombinations. *Folia Biol. (Kraków)* **57**: 149-156.
- KAJTOCH Ł. 2011. Conservation genetics of xerothermic beetles in Europe: the case of *Centricnemus leucogrammus*. *J. Ins. Conserv.* **15**: 787-797.
- KAJTOCH Ł., KOROTYAEV B., LACHOWSKA-CIERLIK D. 2012. Genetic distinctness of parthenogenetic forms of European *Polydrusus* weevils of the subgenus *Scythodrusus*. *Ins. Sci.* **19**: 183-194.
- KAJTOCH Ł., KUBISZ D., GUTOWSKI J.M., BABIK W. 2014. Evolutionary units of *Coraebus elatus* (Coleoptera: Buprestidae) in central and eastern Europe – implications for origin and conservation. *Ins. Conserv. Diver.* **7**: 41-54.
- KAJTOCH Ł., KUBISZ D., LACHOWSKA-CIERLIK D., MAZUR M.A. 2013. Conservation genetics of endangered leaf-beetle *Cheilotoma musciformis* populations in Poland. *J. Ins. Conserv.* **17**: 67-77.
- KAJTOCH Ł., MAZUR M., KUBISZ D., MAZUR M.A., BABIK W. 2014. Low effective population sizes and limited connectivity in a xerothermic beetle: Implications for conservation of an endangered habitat. *Anim. Conserv.* DOI:10.1111/acv.12110.
- KITSON J.J.N., WARREN B.H., VINCENT FLORENS F.B., BAIDER C., STRASBERG D., EMERSON B.C. 2013. Molecular characterization of trophic ecology within an island radiation of insect herbivores (Curculionidae: Entiminae: Cratopus). *Mol. Eco.* **22**: 5441-5455.
- KRESS W.J., ERICKSON D.L. 2007. A Two-Locus Global DNA Barcode for Land Plants: The Coding *rbcL* Gene Complements the Non-Coding *trnH-psbA* Spacer Region. *PLOS ONE* **2** e508.
- KUBISZ D., KAJTOCH Ł., MAZUR M.A., RIZUN V. 2012a. Molecular barcoding for central-eastern European *Crioceris* leaf-beetles (Coleoptera: Chrysomelidae). *Cent. Eur. J. Biol.* **7**: 69-76.
- KUBISZ D., KAJTOCH Ł., MAZUR M.A., LIS A., HOLECOVÁ M. 2012b. Conservation genetics of highly isolated populations of xerothermic *Crioceris quatuordecimpunctata* (Coleoptera: Chrysomelidae). *Inv. Biol.* **131**: 333-344.
- LACHOWSKA D., ROZEK M., KAJTOCH Ł., HOLECOVA M. 2006. Cytogenetic differences between *Peritelus familiaris* and *Centricnemus leucogrammus* (Coleoptera: Curculionidae: Entiminae: Peritelini). *Eur. J. Entomol.* **103**: 687-690.
- LIANA A. 1987. Orthoptera of xerothermic habitats in Poland and their origin In: Baccetti, B (ed) *Evolutionary biology of orthopteroid insects* Ellis Horwood, Chichester, pp 342-346.
- MARRERO P., OLIVEIRA P., NOGALES M. 2004. Diet of the endemic Madeira Laurel Pigeon *Columba trocaz* in agricultural and forest areas: implications for conservation. *Bird Conserv. Int.* **14**: 165-172.
- MARTIN M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*, North America, **17**. <http://journal.embnet.org/index.php/embnet-journal/article/view/200>.
- MARYAŃSKA-NADACHOWSKA A., LACHOWSKA-CIERLIK D., DROSOPoulos S., KAJTOCH Ł., KUZNETSOVA V.G. 2010. Molecular phylogeny of the Mediterranean species of *Philaenus* (Hemiptera: Auchenorrhyncha: Aphrophoridae) using mitochondrial and nuclear DNA sequences. *Syst. Entomol.* **25**: 318-328.
- MATHESON C.D., MULLER G.C., JUNNILA A., VERNON K., HAUSMANN A., MILLER M.A., GREENBLATT C., SCHLEIN

- Y. 2008. A PCR method for detection of plant meals from the guts of insects. *Org. Diver. Evol.* **7**: 294-303.
- 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. 2000. Ochrona owadów siedlisk kserotermicznych Polski [Insect conservation in xerothermic habitats in Poland]. *Wiad. Entomol.* **2**: 129-137.
- MICHALIK S., ZARZYCKI K. 1995. Management of xerothermic grasslands in Poland: botanical approach. *Colloquium Phytosociologiae* **24**: 881-895.
- NAVARRO S.P., JURADO-RIVERA J.A., GOMEZ-ZURITA J., LYAL C.H.C., VOGLER A.P. 2010. DNA profiling of host-herbivore interactions in tropical forests. *Ecol. Entomol.* **35**: 18-32.
- POMPANON F., DEAGLE B.E., SYMONDSON W.O.C., BROWN D.S., JARMAN S.D., TABERLET P. 2012. Who is eating what: diet assessment using next generation sequencing. *Mol. Ecol.* **21**: 1931-1950.
- RYDER O.A. 1986. Species conservation and systematics: the dilemma of the subspecies *TREE* **1**: 9-10.
- SANDHOLM H.A., PRICE R.D. 1962. Field observations on the nectar feeding habits of some Minnesota mosquitoes. *Mosquito News* **22**: 846-849.
- SHOKRALLA S., SPALL J., GIBSON J.F., HAJIBABAEI M. 2012. Next-generation DNA sequencing technologies for environmental DNA research. *Mol. Ecology* **21**: 1794-1805.
- SOININEN E.M., VALENTINI A., COISSAC E., MIQUEL C., GIELLY L., BROCHMANN C., BRYSTING A.K., ŠRNŠTEBŔ J.H., IMS R.A., YOCOZ N.G., 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 mixture. *Front. Zool.* **6**: 16.
- STAUDACHER K., WALLINGER C., SCHALLHART N., TRAUGOTT M. 2011. Detecting ingested plant DNA in soil-living insect larvae. *Soil Biol. Biochem.* **43**: 346-350.
- TABERLET P., GIELLY L., PAUTOU G., BOUVET J. 1991. Universal primers for amplification of three noncoding regions of chloroplast DNA. *Plant Mol. Biol.* **17**: 1105-1109.
- TABERLET P., COISSAC E., POMPANON F., GIELLY L., MIQUEL C., VALENTINI A., VERMAT T., CORTIER G., BROCHMANN C., WILLERSLEV E. 2007. Power and limitations of the chloroplast *trnL* (UAA) intron for plant DNA barcoding. *Nuc. Acids. Res.* **35**: e14.
- TABERLET P., COISSAC E., POMPANON F., BROCHMANN C., WILLERSLEV E. 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecology* **21**: 2045-2050.
- VALENTINI A., POMPANON F., TABERLET P. 2009a. DNA Barcoding for ecologists. *TREE* **24**: 110-117.
- VALENTINI A., MIQUEL C., NAWAZ M.A., BELLEMAIN E., COISSAC E., GIELLY L., POMPANON F., CRUAUD C., NASCETTI G., WINKER P., SWENSON J.E., TABERLET P. 2009b. New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the *trnL* approach. *Mol. Ecology Res.* **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. *Mol. Ecology Res.* **13**: 75-83.
- WAŚOWSKA M. 2006. Chrysomelid communities (Chrysomelidae, Coleoptera) of xerothermic grasslands (*Inuletum ensifoliae*) in the Wyzyna Miechowska Uplands (Central Poland). *Biologia, Bratislava* **61**: 565-572.
- YU J., XUE J.H., ZHOU S.L. 2011. New universal matK primers for DNA barcoding angiosperms. *J. System. Evol.* **49**: 176-181.