Integrative taxonomy of genus *Electrogena* (Ephemeroptera: Heptageniidae): the role of innovative morphological analyses for species delimitation

MAREK POLÁŠEK*,[1], ROMAN J. GODUNKO[2,3], SEREINA RUTSCHMANN[4], MAREK SVITOK[5,6], MILAN NOVIKMEC[5] & SVĚTLANA ZAHRÁDKOVÁ[1]

[1] Masaryk University, Department of Botany and Zoology, Kotlářská 2, CZ-61137 Brno, Czech Republic; Marek Polášek * [m.polasek@mail.muni.cz]; Světlana Zahrádková [zahr@sci.muni.cz] — [2] Czech Academy of Sciences, Biology Centre, Institute of Entomology, Branišovská 31, CZ-37005 České Budějovice, Czech Republic; Roman J. Godunko [godunko@seznam.cz] — [3] State Museum of Natural History, National Academy of Sciences of Ukraine, Teatralna 18, UA-79008 Lviv, Ukraine — [4] Department of Biochemistry, Genetics and Immunology, University of Vigo, 36310, Vigo, Spain; Sereina Rutschmann [sereina.rutschmann@gmail.com] — [5] Faculty of Ecology and Environmental Sciences, Technical University in Zvolen, T. G. Masaryka 24, 96053 Zvolen, Slovakia; Marek Svitok [svitok@tuzvo.sk]; Milan Novikmec [novikmec@tuzvo.sk] — [6] Department of Ecosystem Biology, Faculty of Science, University of South Bohemia, Branišovská 1760, CZ-37005 České Budějovice, Czech Republic — * Corresponding author

Abstract. The easy, fast and correct identification of species is a crucial aspect of biology and its applications, such as biomonitoring and nature conservation. One of the groups that are common but not easily to identify are mayflies at the larval stage. In recent years, many attempts to species identification using modern and non-traditional methods have been made. Two different approaches are used in most cases: i) DNA taxonomy and ii) modern image processing and classification. In this study, we combined both to describe the intrageneric genetical and morphological variability of the Central European representatives of the genus *Electrogena* Zurwerra & Tomka, 1985 – one of several mayfly groups with unclear taxonomy. We compared the classical morphometric method for *Electrogena* species identification with non-traditional Fourier shape descriptor analysis. In particular, we used Linear Discriminant Analysis and Fourier analysis to distinguish among operational taxonomic units defined by generalised mixed Yule-coalescent (GMYC) approach based on cytochrome *c* oxidase subunit I (*cox*1) barcoding gene. Our results demonstrate that the use of modern morphometric methods can significantly improve the proportion of correctly identified individuals to species level. Moreover, the Fourier shape descriptor based analysis revealed with a remarkably low error rate the geographically separated sub-species within the genus *Electrogena*. Our findings show the possibility of computer-aided mayfly (and possibly other insect orders) taxa identification based on selected body part shapes. These approaches might significantly improve routine invertebrate identification and comparability of identification results across different countries and/or among research teams. Better identification can in turn lead to higher robustness of metacommunity studies, ecological status assessment and other science and practice targeted studies based on invertebrate sampling and identification.

Key words. Ephemeroptera, barcoding, GMYC, geographic distribution, morphometry, Fourier outline analysis.

1. Introduction

Mayflies are hemimetabolous freshwater insects represented by more than 3000 described species worldwide (BARBER-JAMES et al. 2008). Due to the general threat of freshwater environments (cf. DUDGEON et al. 2006) and the importance of mayflies as indicators of climatic and anthropogenic changes (cf. BRITTAIN 2008), the knowledge of alpha taxonomy and diversity of mayflies (and aquatic insects in general) is very important. Routine identification of mayfly taxa for purposes of monitoring their distribution and water quality assessment is mostly based on the larval material. However, the discrimination of some mayfly species

based on larval stages remains difficult and in some cases impossible.

One group of mayflies with ambiguous taxonomy is the genus *Electrogena* Zurwerra & Tomka, 1985. At present, the genus *Electrogena* includes about 40 species from the Palearctic and Oriental region (KLUGE 2004). In total, 22 species are known to occur in Europe (excluding Turkey and the Caucasus), and four species in Central Europe – *Electrogena affinis* (Eaton, 1885), *Electrogena lateralis* (Curtis, 1834), *Electrogena ujhelyii* (Sowa, 1981), and *Electrogena quadrilineata* (Landa, 1970). The status of *Electrogena samalorum* (Landa, 1982) is considered as a junior synonym of *E. ujhelyii*. The synonymy was established by ZURWERRA & TOMKA (1986) with reference to the personal communication of Dietrich Braasch, but no other details or comments were given. Later, the synonymy was also supported by KLUGE (2004), but also without any details.

Most of the European species of the genus *Electrogena* were originally described as representatives of other Heptageniidae genera belonging to the subfamily Ecdyonurinae, mostly as species of the genus *Ecdyonurus* Eaton, 1868 or *Heptagenia* Walsh, 1863. The genus *Electrogena* was designated on the basis of the former *Ecdyonurus lateralis* group using the electrophoretic separation of allozymes – biochemical products of several enzyme-coding genes (ZURWERRA & TOMKA 1985). In the studies by ZURWERRA & TOMKA (1985) and ZURWERRA et al. (1987), the position of the genus *Electrogena* among the other valid genera and the relationships within the genus were established. A recent study by YANAI et al. (2017) provides a revised generic concept of the Ecdyonurinae mayflies, revealing the former species *Electrogena zebrata* as belonging to the newly described genus *Anapos* Yanai & Sartori gen.n. (2017). Further YANAI et al. (2017) investigate the phylogenetic relationships among seven *Electrogena* species based on morphological and molecular characters. Most recently, the study by WAGNER et al. (2017) newly described the species *Electrogena brulini* Wagner, 2017 focusing on Alpine representatives of the genus *Electrogena*. While these previous studies have contributed to the consolidation of the generic concepts within Ecdyonurinae, much work remains to be undertaken to establish the relationships within the genus *Electrogena*, focusing on more population genetic levels in order to identify species and their population structures.

The discrimination of Central European *Electrogena* species remains an open quest. Despite several studies on larval morphology and identification of mayflies including *Electrogena* species (SCHOENEMUND 1930; LANDA 1969), the discriminating characters were not unified. Significant contributions to the identification and discrimination of *Electrogena* species based on a unified set of morphological characters of larvae have been carried out by Belfiore and co-workers (i.e., BELFIORE 1994, 1995, 1996, 1997; BELFIORE & DESIO 1995; BELFIORE et al. 1997, 1999, 2000; BELFIORE & SARTORI 1999; RIGHETTI & BELFIORE 2000). BELFIORE (1994) established the system of well-defined larval characters divided into three groups, including meristic, ratio and descriptional/qualitative characters. Each character is coded according to the type of character and relevant body part. These established morphological characters have been used for several more recent taxonomic works like description of *Electrogena braaschi* (Sowa, 1984) (GODUNKO 2000; GODUNKO et al. 2002), redescription of *E. quadrilineata* with notes on related species (KŁONOWSKA-OLEJNIK 2004), description of *Electrogena gibedede* Sroka & Godunko, 2012 from Caucasus (SROKA & GODUNKO 2012), description of larvae from *Ecdyonurus venosus* group (HAYBACH 1999) and description of new species *E. brulini* (WAGNER et al. 2017). Several morphological characters proposed in HAYBACH (1999) were adopted for this study.
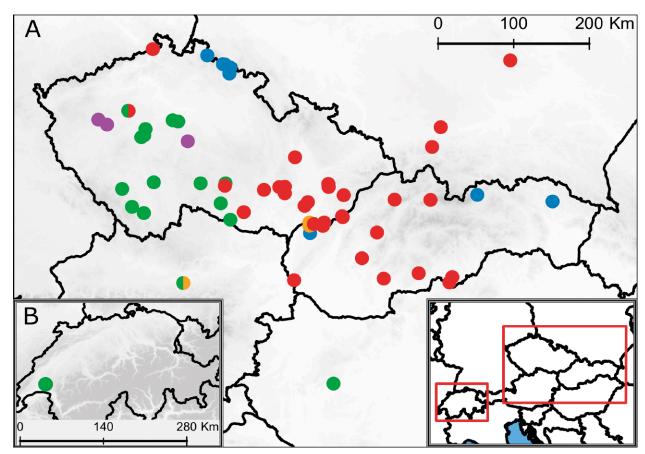
Based on Belfiore's work we assume that the application of the set of qualitative and quantitative larval characters mentioned above should be sufficient for distinguishing the Central European *Electrogena* species. However, most of the considered studies have been dealing with Alpine or Italian *Electrogena* species. Therefore a comparison of all Central European species (including *E. quadrilineata*) has not been carried out yet. And despite the opinion that *E. samalorum* is a junior synonym of *E. ujhelyii*, the taxonomic position of *E. samalorum* remains unclear: the high intraspecific variability of *E. ujhelyii* discussed in BELFIORE (1996) could indicate that *E. ujhelyii* is a complex of cryptic species. Moreover, the autecological preferences of both species are stated differently (SOWA 1981; LANDA & SOLDÁN 1982) which also suggests presence of cryptic species.

In the present study, we examined the morphological and genetic differences among Central European *Electrogena* representatives. To delineate the lineages based on the barcoding gene, we applied the generalised mixed Yule-coalescent (GMYC) approach (FUJISAWA & BARRACLOUGH 2013). Further, we compared the larval morphological differences among the operational taxonomic units (OTUs) revealed by molecular analysis (GMYC OTUs). We combined multidimensional analyses on a set of morphological characters (based on Belfiore's system) and Fourier outline analysis of selected body parts. We discuss the cryptic diversity and geographic distribution of selected GMYC OTUs, and provide a revision of larval morphological differences among *Electrogena* species. A particular emphasis has been placed on *E. ujhelyii* and possible *E. samalorum*.

2. Material and methods

2.1. Field sampling

Larvae of Central European *Electrogena* species were collected between 2006 and 2014, using the kick-sampling method and hand-net sampling. We focused on material from the Czech Republic and Slovakia because this

Fig. 1. Map of sampling sites with GMYC OTUs occurrence. **A**: Map of Central Europe covering the majority of sampled sites. **B**: Map of Switzerland with one sampled site. Position of both figured regions is highlighted by red rectangle in the bottom-right overview map. Coloured dots indicate occurrence of different GMYC OTUs, bicoloured dots indicate two different GMYC OTUs occurring at one locality. — **Colours**: violet – affinis, red – eastern ujhelyii, blue – lateralis, orange – quadrilineata, green – western ujhelyii.

is the area of occurrence of the putative cryptic species. The rest of the collected material came from Hungary (one site was sampled – a brook near Aszófő village, type locality of *E. ujhelyii*), Austria, Poland, and Switzerland (see Fig. 1). The collected material was preserved in 99% ethanol and stored at 4°C. Altogether, we used 112 individuals for morphological analyses, of which 69 individuals were used for the molecular analyses (for details see Supplement Table S1).

2.2. DNA extraction, amplification, and sequencing

We selected one to two specimens per locality for the molecular analysis. Ideally, two legs (if possible, the right fore and middle leg) were used for DNA extraction. The DNA was extracted and purified using the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany). We sequenced the 658-bp fragment of the mitochondrial cytochrome *c* oxidase subunit I (*cox*1) gene, using the primer pair LCO1490 and HCO2198 (Folmer et al. 1994). The polymerase chain reaction (PCR) included an initial denaturation (2 min at 94°C) followed by 37 thermocycles of denaturation (30 s at 94°C), annealing (45 s at 45°C), extension (50 s at 72°C), and a final exten-

sion phase (7 min at 72°C). The length of the amplified products was pre-checked via agarose gel electrophoresis. All amplified PCR products were purified using the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany). The purified PCR products were used as templates for cycle sequencing reactions with BigDye® Terminator v3.1 (Applied Biosystems), following the manufacturer´s instructions. Forward and reverse sequences were read with a capillary sequencer (Applied Biosystems) at the Department of Botany and Zoology, Masaryk University, Brno, Czech Republic. All sequences were assembled and edited with Sequencer® v5.3 (Gene Codes Corporation, Ann Arbor, MI USA).

2.3. DNA taxonomy

Genetic species delineation was carried out by combining newly sequenced *cox*1 data with 25 published *Electrogena* sequences available from GenBank (Cardoni et al. 2015; Gattolliat et al. 2015; Wakimura et al. 2016). As outgroup we used *Baetis rhodani* (GenBank acc. no KF438126) *sensu* Rutschmann et al. (2017b). The *cox*1-sequences were aligned using MAFFT v7.221 (Katoh & Standley 2013). To check for stop codons and the occurrence of indels, we translated the nucleotide sequence

alignment into an amino acid alignment using MEGA 5 (TAMURA et al. 2011). Identical haplotypes were removed using the Perl script collapsetypes_v4.5.pl (CHESTERS 2013). The sequence alignment was partitioned into codon positions (first codon position, second codon position, and third codon position). The best-fitting model of molecular evolution was selected for each codon position separately with jModelTest v2.1 (GUINDON & GASCUEL 2003; DARRIBA et al. 2012) using the Bayesian information criterion (BIC). The selected available models were: GTR + I for codon position one, HKY for codon position two, and GTR + Γ for codon position three. Since this partitioning-scheme resulted in very low effective sample sizes (ESS) for preliminary tree inferences, we used the codon-specific model of sequence evolution HKY_{112} + CP_{112} + $Γ_{112}$ (SRD60, SHAPIRO et al. 2006) for the final analysis. An ultrametric gene tree was calculated using a relaxed molecular clock and a coalescent prior *sensu* MONAGHAN et al. (2009) in BEAST v2.3.1 (BOUCKAERT et al. 2014). For the tree reconstruction, we conducted six independent runs with 30 million generations each. We used Tracer v1.7.1 (RAMBAUT et al. 2018) to assess convergence between runs by inspecting marginal densities and ESS values, and estimated the number of burn-in generations for each run. All runs resulted in trees with very similar topologies (i.e., each putative GMYC species was recovered as monophyletic clade, except for one run for which *E. lateralis* from Switzerland and France, was not recovered as monophyletic clade, see Results). The runs were combined in LogCombiner v2.3 (RAMBAUT & DRUMMOND 2015), after removing the first 10% (i.e., 3000 trees) of each run as burn-in. Maximum clade credibility trees were inferred with TreeAnnotator v2.3 (RAMBAUT & DRUMMOND 2015). A single threshold GMYC analysis was performed on the ultrametric gene tree, using the Splits package (http://r-forge.r-project.org/projects/splits/) for R v3 (R Core Team, 2017).

2.4. Morphology and morphometry

For 13 Palearctic *Electrogena* species, we obtained a set of 13 numerical and four qualitative characters based on previous work by Belfiore (i.e., BELFIORE 1981, 1994, 1995, 1996; BELFIORE & DESIO 1995; BELFIORE et al. 1997, 1999, 2000; BELFIORE & SARTORI 1999; RIGHETTI & BELFIORE 2000; KŁONOWSKA-OLEJNIK 2004). The minimum, mean and maximum of numerical characters were taken from mentioned papers when available; the status of qualitative characters was noted. When a certain numerical character or status of qualitative character of a single species differed between two (or more) papers, we used the information from the more recent publication.

A set of several individuals from each locality were chosen regarding the condition of the preserved material. Only individuals with minimal body damage were chosen for morphological analyses. The individuals included in the morphological analyses were selected with a particular emphasis on maximal morphological vari-

ability within a single locality – only specimens of different size, age and sex classes (without young larvae) from one locality were included. At least one individual from each locality was analysed: the voucher specimen used for the molecular study. The maximum number of examined individuals from one locality was five. Each specimen was labelled by a unique identification code, stored separately in a plastic phial, and processed as described below. The photographs of gill plates in water medium were taken for preserving the exact shape of the soft parts, which are usually deformed in the non-water mounting medium. The whole habitus and details of the head were also captured. The mouthparts, in some cases also the head capsule, 1st, 4th and 7th tracheal gill plate, and fore, mid and hind legs were mounted into Liquido-Faure on microslides.

A binocular stereomicroscope Olympus SZX9 with mounted digital camera Olympus C9090 was used for preparing the microslides and capturing the photographic material. For more detailed imaging of some morphological characters, the microscope Olympus BX41 with mounted digital camera was used. The images were pre-processed using the software QuickPHOTO MICRO v3.1 (www.promicra.cz), and further image adjustment was done with the GNU Image Manipulation Program v2.8 (GIMP; www.gimp.org).

2.4.1. Morphological characters

The set of 16 quantitative numerical, seven ratio and six qualitative characters were defined, mostly based on those described in Belfiore's articles mentioned above. Binary matrices for each character state were used for coding presence or absence of certain qualitative characters. The presence (in data matrix coded as 1) or absence (coded as 0) of each binary variable (highlighted in bold) was marked. The description of morphological characters and respective indexes are listed in the Appendix.

2.4.2. Morphometric data analyses

We used the matrix with character values (continuous for quantitative, binary for qualitative characters) of each examined individual (see Supplement Table S1) for morphometric analyses. Each individual was assigned to the OTU according to the species assignment obtained *via* GMYC approach (GMYC OTUs). The minimum, mean and maximum values for each numerical character were calculated. To illustrate the differences between given GMYC OTUs, the boxplots (the 25th and 75th percentiles) of numerical characters and histograms of qualitative characters represented as dummy variables presence/absence frequencies are used for visual presentation of differences between them. The significance of each numerical character was expressed using Kruskal-Wallis H test between groups (GMYC OTUs).

To distinguish i) among individuals from all available GMYC OTUs and ii) between the closely related individuals within *E*. cf. *ujhelyii* (e.g., *E. ujhelyii* sensu stricto

and possible *E. samalorum* together), we performed a Linear Discriminant Analysis (LDA). For this analysis, the missing values in the character data matrix (in the case when the value of certain characters was not possible to count due to missing body part or body part damage) were estimated using factorial analysis for mixed data (FAMD) method (AUDIGIER et al. 2014). Before LDA analysis, we reduced the number of morphological variables to remove the collinearity and improve the separation power of variables. To select the most informative variables, we used the proportion of between-group and within-group variance of each variable as a criterion. We chose two different sets of variables, discriminating all GMYC OTUs (full dataset) and *E.* cf. *ujhelyii* GMYC OTUs (reduced dataset) respectively.

Additionally, we presented the equation for distinguishing between closely related *E.* cf. *ujhelyii* individuals using selected morphological characters. The success rate of identifying both *Electrogena* and *E.* cf. *ujhelyii* individuals using LDA analyses of numerical characters is presented as leave-one-out cross-validation success rates.

2.4.3. Outline analysis

Images of the 1st, 4th and 7th tracheal gill, glossa, labrum and metafemora were captured as mentioned above. The images were trimmed on 1000×1000 px image and adjusted, thus the target body part was approximately in the middle of the picture. All images were manually transformed to binary black and white images using GIMP v2.8. The outlines of these shapes were extracted, and the elliptic Fourier analysis was performed (for details see CLAUDE 2008). Shapes were approximated by 32 harmonics. Principal component analysis (PCA) on harmonic coefficients was calculated to describe the highest shape variability among individuals for each character. Linear Discrimination Analysis was performed on PCA scores to test the possibility of discriminating the GMYC OTUs using morphological shapes. To compare the classical morphometry methods and modern image-based analyses, we used the same data partitioning as for the morphometric dataset. We performed all analyses on the complete dataset including all GMYC OTUs and the reduced dataset including only *E.* cf. *ujhelyii*. The success rate of identification *Electrogena* OTUs using image-based LDA analyses is presented as leave-one-out cross-validation confusion matrix. All operations were performed using the package Momocs (BONHOMME et al. 2014) in R.

3. Results

3.1. GMYC analysis

In total we found 65 unique haplotypes within the 94 sequences. The GMYC model was significant (likelihood ratio: 14.03, P = 0.003). The model delineated twelve putative species, composed of eight distinct clusters and four singletons (Fig. 2). The 95% confidence interval (CI, defined as 2 log likelihood units) ranged from ten to 13 species and six to nine clusters. All following results are based on the final model as presented in Fig. 2.

We detected ten putative GMYC species of the genus *Electrogena*. In detail, the model detected the species *E. affinis*, *E. grandiae* (including three unassigned specimens indicated as *E.* sp.), and *E. fallax*, each as individual GMYC clusters; *E. quadrilineata* and one unassigned *Electrogena* specimen from Japan (*E.* sp.) as singleton. The morphospecies *Anapos zebratus* (formerly *E. zebrata*) formed two putative GMYC species, including a singleton from Corsica, and a cluster of six haplotypes from Sardinia.

The two morphospecies *E. lateralis* and *E.* cf. *ujhelyii* were each divided into two putative GMYC species. In case of *E. lateralis*, one putative GMYC species included individuals from France and Switzerland, the other one comprised individuals from the Czech Republic and Slovakia. *Electrogena* cf. *ujhelyii* consisted of two clusters with a geographic contact line in the middle of the Czech Republic (Fig. 1). The first putative GMYC species comprised specimens from Switzerland, Austria, Hungary, and Czech Republic (i.e., Bohemia – western part of the Czech Republic). The second putative GMYC species included specimens from Czech Republic (i.e., Moravia – eastern part of the Czech Republic), Slovakia, and Poland. Interestingly, at one locality both putative GMYC species from *E.* cf. *ujhelyii* group occurred together (locality Klíčava river, Lány, North-West of the Czech Republic).

3.2. Morphometry

The overview of available information on morphological characters taken from published data is listed in Supplement Table S2. Data were available for species *E. affinis*; *Electrogena antalyensis* (Braasch & Kazanci, 1986); *Electrogena calabra* Belfiore, 1995; *Electrogena fallax* (Hagen, 1864); *Electrogena galileae* (Demoulin, 1973); *Electrogena grandiae* (Belfiore, 1981); *Electrogena gridelii* (Grandi, 1953); *Electrogena hyblaea* Belfiore, 1994; *E. lateralis*; *Electrogena lunaris* Belfiore & Scillitani, 1997; *Electrogena malickyi* (Braasch, 1983); *E. quadrilineata* and *E. ujhelyii*.

Based on the results of the GMYC analysis, we separated the sampled material available for morphological analyses into five GMYC OTUs: **affinis**, **lateralis**, **quadrilineata**, **eastern ujhelyii** and **western ujhelyii**. These categories are used in all subsequent morphological analyses.

The minimum, mean and maximum of 23 numerical (discrete and ratio) characters selected for this study were calculated from a morphological dataset of 112 individuals classified in five GMYC OTUs (see Supplement Table S3). The frequencies of qualitative traits occurrences are also listed in Supplementary file (Table S4).
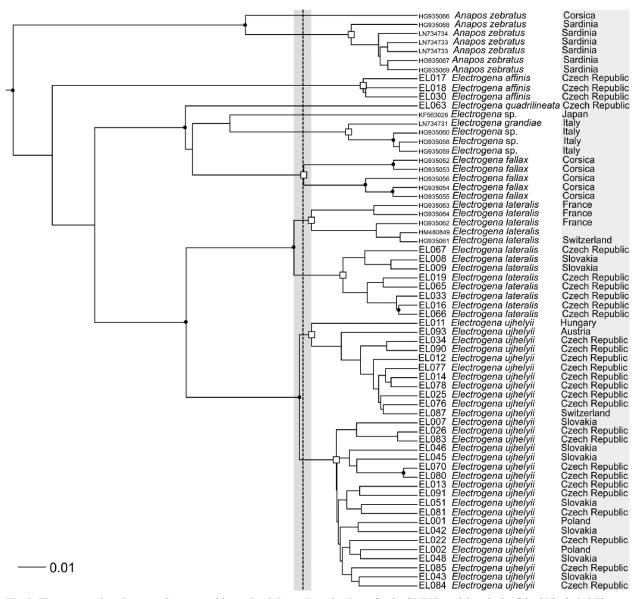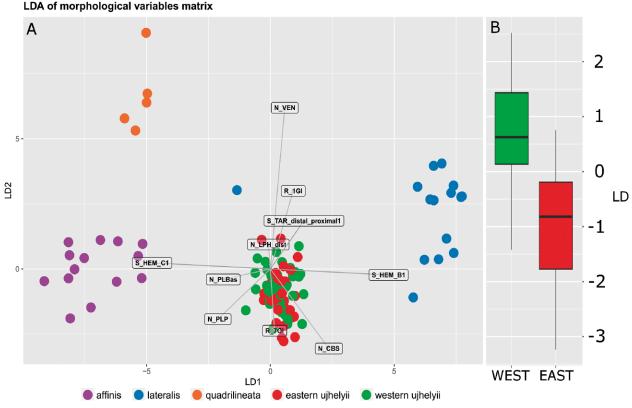
Fig. 2. The gene tree based on cytochrome *c* oxidase subunit I (*cox*1) used as input for the GMYC model analysis of the 93 included *Electrogena* specimens. Sequence clusters corresponding to single GMYC species are indicated by open squares. Filled circles indicate supported nodes (Bayesian posterior probability ≥ 0.95). Terminal labels indicate voucher identification number, morphological assignment, and sampling site origin (see Supplement Table S1). Bigger font labels indicate sequences obtained from sampled material, smaller font labels indicate sequences obtained from GenBank. The dotted line indicates the point of maximum-likelihood fit of the single-threshold GMYC model. The grey shading corresponds to the 95% confidence intervals of the GMYC analysis. Scale bar indicates substitutions per site.

Additionally, the boxplots for numerical characters and frequency histograms for qualitative characters are presented in Supplement Figs. S1, S2 and S3, respectively. All numerical characters, except N_LBB and R_LBR, differ significantly (according to Kruskal-Wallis H test; $p < 0.05$) for at least one OTU; all numerical characters except N_TCB, N_MPB_L, N_MPB_R and R_4GI have p-values lower than 0.001.

The comparison of basic metrics (minimum, mean and maximum of numerical characters, and presence/absence of character state of qualitative characters) of morphological characters defined by Belfiore (see above) revealed relatively high dissimilarity between our findings and published results. The minimum, mean and maximum of characters N_CBS, N_TCB, N_OUT and

N_PLP, and minimum-maximum distance of R_LBR, R_1GI and R_7GI characters were higher in published data than in our observation for all examined species. The other compared numerical characters were more or less similar. The divergences to published data were more common for *E.* cf. *ujhelyii*. The observed states of qualitative characters were in agreement with published data (see Supplement Table S2).

The noncollinear characters with highest separation power for discrimination of all selected GMYC OTUs (see Appendix for abbreviations explanation; the characters are ordered from those with highest separation power to those with lowest separation power) are **S_HEM_C**, **S_HEM_B**, **N_VEN**, **S_TAR_distal_proximal**, **N_PLP**, **R_7GI**, **N_CBS**, **N_PLBas**, **R_1GI**, **N_LPH_**

Fig. 3. Linear Discriminant Analysis of 10 selected morphological characters. **A**: The LDA biplot is based on the complete dataset. The percentage separation achieved by first two linear discriminant functions is 71.8% and 17.3%, respectively. **B**: The boxplots of discriminant scores (LD) given by discriminant equation for all individuals within eastern and western ujhelyii GMYC OTUs. The equation was adjusted, so the separation value is equal to zero.

dist. The LDA biplot for two first linear discriminant axes is presented in Fig. 3A. All valid Central European species were distinguished using selected morphological characters. The success rate of species identification was 100% for *E. affinis*, *E.* cf. *ujhelyii* (eastern and western ujhelyii together) and *E. quadrilineata*, and 93.3% for *E. lateralis*. The percentage of correctly identified individuals within *E.* cf. *ujhelyii* was 51.6% for eastern ujhelyii and 78.7% for western ujhelyii respectively.
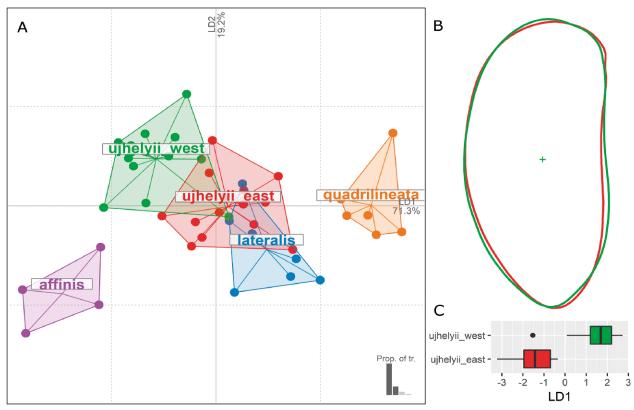
The results of LDA on reduced (eastern and western ujhelyii GMYC OTUs) dataset are presented as an equation for discriminant score (DS) calculation:

$$DS = 0.561*\textbf{N\_MPB\_R} + (-0{,}224*\textbf{N\_LPH\_dist}) + (-2.103*\textbf{R\_7GI}) + (0.388*\textbf{N\_CLW}) + (-0.845*\textbf{R\_GLA}) + (1.785*\textbf{S\_TAR\_distal}) + (0.517*\textbf{N\_HFF}) + (0.230*\textbf{S\_MFE\_symetric}) + (-0.043*\textbf{N\_PLP}) + (-0.136*\textbf{N\_TCB}) + 7.2$$

Values of DS above zero indicate western ujhelyii and values below zero indicate eastern ujhelyii GMYC OTU. The characters included in the equation (highlighted in bold) are ordered from those with highest separation power (N_MPB_R) to those with lowest separation power (N_TCB). The distributions of discriminant scores are presented as boxplots in Fig. 3B. The success rate of identified individuals within *E.* cf. *ujhelyii* was 67.7% and 80.6% for eastern and western ujhelyii GMYC OTUs respectively.

3.3. Elliptic Fourier outline analysis

Altogether, 272 images of different body parts were used for Fourier outline analyses: 45 images of 1st gill plate, 47 of 4th gill plate, 42 of 7th gill plate, 47 of glossae, 39 of labrum, and 52 of metafemur. When trying to distinguish among all selected GMYC OTUs or between closely related eastern and western ujhelyii units, the most informative shapes were the outlines of gill plates, especially the 7th one (Fig. 6A,B). Based solely on this outline shape, all (except one individual of *E. lateralis*) valid Central European *Electrogena* species were correctly identified. The identification success ratio of eastern and western ujhelyii GMYC OTUs was also unexpectedly high (approx. 90% of correctly identified individuals). This percentage is above a level of eastern and western ujhelyii GMYC OTUs correct identification using the morphological characters (approx. 75% of the correctly identified individuals, see above). Average shapes of this particularly informative body part for both GMYC OTUs are presented in Fig. 6B. The other gill plate outlines (first and fourth gill plate) were less informative. The identification success rate of valid Central European *Electrogena* species was 77.8% and 80.9% (for first and fourth gill plate respectively), whereby the eastern and western ujhelyii GMYC OTUs were correctly identified in 67.9% and 74.1% (Figs. 4A,B, 5A,B). This identifying success rate is surprisingly high, particularly in case

Fig. 4. Fourier outline analysis on 1st gill plate. **A**: LDA on PCA coefficients of 1st gill plate shape Fourier harmonics. Biplot of LD scores for all GMYC OTUs on left, with first LD axis covering 71.3% of shape variability. **B**: Average shapes of 1st gill plate for both closely related GMYC OTUs (eastern and western ujhelyii). **C**: Boxplots of LD scores for eastern and western ujhelyii GMYC OTUs.
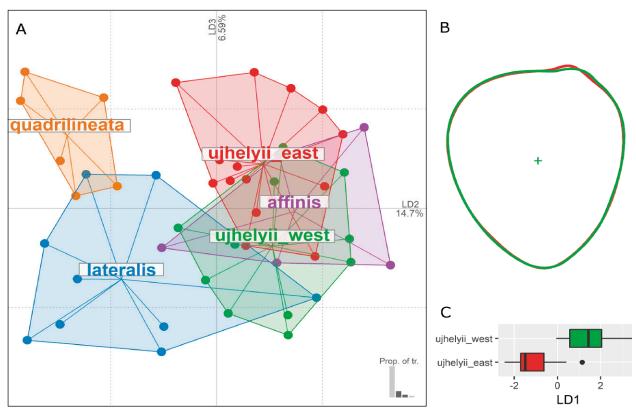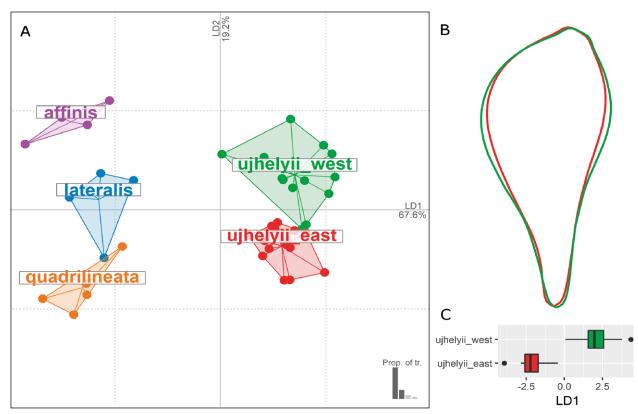

Fig. 5. Fourier outline analysis on 4th gill plate. **A**: LDA on PCA coefficients of 4th gill plate shape Fourier harmonics. Biplot of LD scores for all GMYC OTUs on left, second and third axis were selected for displaying the biplot. First axis represented the differences between shapes of 4th gill plate of affinis GMYC OTU and the other groups. Therefore, the first axis was not very informative in order to distinguish among all analyzed GMYC OTUs. **B**: Average shapes of 4th gill plate for both closely related GMYC OTUs (eastern and western ujhelyii). **C**: Boxplots of LD scores for eastern and western ujhelyii GMYC OTUs.

Fig. 6. Fourier outline analysis on 7th gill plate. **A**: LDA on PCA coefficients of 7th gill plate shape Fourier harmonics. Biplot of LD scores for all GMYC OTUs on left, with first LD axis covering 67.6% of shape variability. **B**: Average shapes of 7th gill plate for both closely related GMYC OTUs (eastern and western ujhelyii). **C**: Boxplots of LD scores for eastern and western ujhelyii GMYC OTUs.

Table 1. Relative success rates (%) of correct OTUs identification by LDA on Fourier outline descriptors using leave-one-out cross-validation. The closely related GMYC OTUs (eastern and western ujhelyii) are evaluated in two ways: i) as one group (*E.* cf. *ujhelyii*) according to the present taxonomical status and ii) separately as two independent groups (east. ujhelyii and west. ujhelyii). In second case (groups = GMYC OTUs) the relative success rates for affinis, lateralis and quadrilineata GMYC OTUs are the same as in the first case (groups = species).

OUTLINE	groups = species					groups = GMYC OTUs		
	E. affinis	*E. lateralis*	*E. quadrilineata*	*E.* cf. *ujhelyii*	**TOTAL 1**	eastern ujhelyii	western ujhelyii	**TOTAL 2**
gill 1	75	57.1	83.3	82.1	**77.8**	57.1	78.6	**68.9**
gill 4	100	33.3	66.7	96.3	**80.9**	73.3	75	**68.1**
gill 7	100	80	100	100	**97.7**	86.7	92.9	**90.7**
glossae	0	40	80	88.2	**76.6**	29.4	41.2	**38.3**
labrum	66.7	40	40	84.6	**71.8**	58.3	42.9	**48.7**
femur	60	66.7	25	97	**82.4**	37.5	44.4	**46.2**

of 4th gill plate which is almost identical in both *E.* cf. *ujhelyii* GMYC OTUs (see Fig. 5B). The identification success rate by use of outline shapes of glossae, labrum and metafemur differed relatively to taxonomical resolution of analysed individuals: whereas all these shapes delimited all valid Central European *Electrogena* spe-

cies (without distinguishing between eastern and western ujhelyii GMYC OTUs) with remarkably high resolution rate (for cross-validation success rates for all shape outlines see Table 1), the same shapes were much less informative when distinguishing also between eastern and western ujhelyii GMYC OTUs.

4. Discussion

4.1. Species diversity

The successful species identification of European *Electrogena* species, combining DNA taxonomy *via* GMYC and morphometric analyses confirmed the broad utility of our approach. In recent years, several studies on mayflies have focused on the use of the barcoding gene *cox1* for species delimitation (e.g., Vuataz et al. 2011; Rutschmann et al. 2014) and the combination of several molecular markers for phylogeographic studies (e.g., Vuataz et al. 2016; Rutschmann et al. 2017a,b).

The GMYC analysis revealed intrageneric diversity for *E. lateralis* and *E. ujhelyii*, including a geographical separation into two distinct putative GMYC species (see Fig. 2). The geographical separation of two *E. lateralis* GMYC species was suggested by Haybach (2008), who defined two subspecies: *E. lateralis lateralis* (Curtis,

1834) from Germany and Ireland, and *E. lateralis concii* (Grandi, 1953) from Italy, Switzerland and Austria; both subspecies have different egg structure. Recent molecular data also evidenced a geographic clustering of *E. lateralis*, including specimens from the type locality in the United Kingdom and from France (YANAI et al. 2017). Although specimens included in our analysis from France and Switzerland were not available for morphological analysis, we assume that the separation of two putative GMYC species corresponds with Haybach's findings: the Alpine populations of *E. lateralis* belong to subspecies *E. lateralis concii*, and the Czech and Slovak population belong to subspecies *E. lateralis lateralis*.

The geographic clustering of the two putative *E. ujhelyii* GMYC species was rather unexpected since there is no previous evidence from the literature. The western *E. ujhelyii* GMYC species distribution could be described as Alpine-Pannonian; the eastern *E. ujhelyii* species distribution could be described as Carpathian. According to the type locality of the species *E. ujhelyii* (Aszófö, Hungary) and *E. samalorum* (Podhoroď, Slovakia), the western ujhelyii GMYC OTU represents the originally described species *E. ujhelyii*, and the eastern ujhelyii GMYC OTU might represent *E. samalorum* or an unknown species.

Our data indicate distinct intraspecific genetic variability for the different *Electrogena* species. While the within-group genetic distances among *E. affinis* specimens were detected as rather high, the *cox*1-sequences of *E. quadrilineata* specimens were recovered as identical haplotype. Notably the analysed *E. quadrilineata* specimens came from three localities (Czech Republic and Austria) and *E. affinis* from three nearby localities in the Czech Republic.

The geographical sampling of the species *E. affinis*, *E. quadrilineata*, *E. grandiae*, and *E. fallax* was rather restricted (i.e., Czech Republic, Italy, and Corsica). All previous molecular work on these species (CARDONI et al. 2015; GATTOLLLIAT et al. 2015; YANAI et al. 2017) included specimens from the same geographic areas, with the exception of YANAI et al. (2017), who included for *E. quadrilineata* specimens from Czech Republic and Austria. Further molecular analyses on a wide geographical scale will be required to confirm the status of these species.

We confirmed the monophyletic clustering of the recently proposed genus *Anapos* Yanai & Sartori, 2017, including *A. zebratus* (formerly classified as *E. zebrata*) (YANAI et al. 2017). The results of our GMYC analysis are in agreement with previous studies, including the existence of two distinct *A. zebratus* lineages (GATTOLLIAT et al. 2015; YANAI et al. 2017).

4.2. Morphology

The observed differences between measured and published values of numerical and qualitative characters illustrate the risks of applying defined numerical morphologic characters. The resulting value of these characters strongly depends on the available equipment (such as preparation tools and microscope) and personal experience. For example, the description of the character N_CBS (a number of comb-shaped bristles on the fore margin of galeo-lacinia) does not define the threshold between comb-shaped bristle and other-type bristles, which are also present (including a number of the transitional types) on the fore margin of galeo-lacinia. This could explain the differences between our own measured and available published values of this character. The other numerical characters which differed significantly were based on small morphological structures (N_TCB) or counts of very fine setae (N_OUT, N_PLP). The differences between our measurements and published values of ratio characters could be caused by the different method of data collecting. While Belfiore and other authors measured the values from hand-made drawings, we used the photographs of microslide mounted body parts. However, these characters are useful when they have extreme values and can be interpreted as presence or absence of a remarkable state of character. For example, the character N_HFF (the number of long hairs on the fore margin of femora) can be used for the identification of *E. affinis* characterised by a high number of these long hairs, whereas the other *Electrogena* species have a few of them (*E. galileae*) or none (rest of *Electrogena* species).

On the other hand, the qualitative characters are quite informative, and the valid Central European *Electrogena* species can be distinguished solely on the basis of characters S_HEM (distinct light spots for *E. affinis* and *E. quadrilineata*, two lighter smudges for *E. lateralis* and unicolour head for *E.* cf. *ujhelyii*) and S_BFE(i) (pointed bristles for *E. affinis* and spatula-like bristles on the fore femora for *E. quadrilineata*). The LDA results confirmed the high potential of qualitative characters, with the character S_HEM being the most informative character for distinguishing Central European *Electrogena* species. Additionally, we can consider the qualitative characters S_TAR (colouration of tarsus) and S_BFE(i) (shape of bristles on upper side of fore femora) as informative as S_HEM because of high intercorrelation among these characters. The success rate of valid Central European *Electrogena* species identification using the LDA on the set of most informative qualitative and quantitative characters was almost 100%. Therefore, we can confirm the usability of the system of qualitative morphological characters for Central European *Electrogena* valid species identification.

The equation for discriminating the eastern and western ujhelyii GMYC OTU derived from LDA results based on the reduced dataset showed a surprisingly high separation power (67.7% and 80.6% cross-validation success rate, respectively). However, the equation was mostly based on characters, whose values often differed from published ones. Some of the other characters (N_HFF and S_TAR_distal) were included because of few atypical specimens with exceptional values of those characters (and therefore easily recognizable in the dataset).

Hence, it is questionable if this equation based partly on problematic morphologic characters is applicable for distinguishing the closely related *E*. cf. *ujhelyii* individuals.

The Fourier outline shape analysis provided the most exciting results. The percentage of correctly identified individuals in the dataset using the most informative outlines depended on the taxonomical resolution of the analysed specimens. When we tried to distinguish among currently taxonomic valid species of *Electrogena* (e.g., eastern and western ujhelyii GMYC OTUs classified as one species – *E*. cf. *ujhelyii*), the observed success rate varied from almost 100% to approx. 80% depending on which body part was assessed. This was comparable to the success rate of the LDA model on a set of ten most informative qualitative and quantitative characters. The main contribution of the Fourier shape descriptors analysis improved the success rate to distinguish the eastern and western ujhelyii GMYC OTUs. In this case, the LDA model correctly identified more than 90% of the specimens using only the 7th gill plate shape. Moreover, the model successfully distinguished between two individuals belonging to different GMYC OTU but occurred at a single locality (Klíčava river, Lány, Czech Republic). Compared to the success rate of LDA based on the ten most informative traditional characters, the Fourier outline analysis was more accurate using only one character. When we take into account the uncertainty of measuring (or obtaining) the morphological characters entering the model (discussed above), the Fourier descriptor based analysis is the only reliable way to distinguish between eastern and western ujhelyii GMYC OTUs.

However, we did not find any character to confidently differentiate the eastern and western ujhelyii GMYC OTU. Therefore, we decided not to make any statement on taxonomy of those geographically separated (GMYC based) putative species. Even though we suppose, that eastern ujhelyii represents the former species *E. samalorum*, further investigation of different life stages (eggs, adults) is needed to reject (or support) the recent synonymy of *E. ujhelyii* and *E. samalorum*. Moreover, the comparison of type material of both later species has been done, but due to the extent of this work, the results cannot be included in this study and will be published soon. Nevertheless, no significant differences between larval or adult type individuals were recognised (Polášek unpubl.).

The comparison of different species (or GMYC species) identification approaches showed that using modern approaches to shape analysis significantly improves the accuracy of species identification. However, the most recent approaches to automated species identification involve deep neural networks or other computer vision techniques for image (i.e., whole body or body part) classification. Computer vision has been used to classify specimens in different genera (Larios et al. 2008), visually easily separable species (Wen & Guyer 2012; Boniecki et al. 2015) or even closely related species and genera (Favret & Sieracki 2016), including a high proportion of correctly identified specimens. In this study

we demonstrated that using a non-traditional approach (such as outline Fourier descriptors) can significantly improve our ability to identify taxonomical units defined by DNA-taxonomy methods, which are not identifiable using comprehensive traditional morphometric methods. Additionally, the high separation power of easily captured qualitative characters, such as colouration of the head, could indicate the high potential of computer vision use for mayfly species identification. Therefore, the DNA-taxonomy-supported set of diagnostic shapes or other image-based data could contribute to the routine species identification, for example as a computer-vision extension of electronic identification keys.

5. Acknowledgements

6. References

Audigier V., Husson F., Josse J. 2016. A principal component method to impute missing values for mixed data. – Advances in Data Analysis and Classification **10**(1): 5−26. doi:10.1007/s11634-014-0195-1

Barber-James H.M., Gattolliat J.-L., Sartori M., Hubbard M.D. 2008. Global diversity of mayflies (Ephemeroptera, Insecta) in freshwater. – Hydrobiologia **595**(1): 339−350. doi:10.1007/s10750-007-9028-y

Belfiore C. 1981. On the Italian species of the *Ecdyonurus lateralis* group (Ephemeroptera, Heptageniidae). – Aquatic Insects **3**(3): 171−178. doi:10.1080/01650428109361059

Belfiore C. 1994. Taxonomic characters for species identification in the genus *Electrogena* Zurwerra and Tomka, with a description of *Electrogena hyblaea* sp. n. from Sicily (Ephemeroptera, Heptageniidae). – Aquatic Insects **16**(4): 193−199. doi:10.1080/01650429409361555

Belfiore C. 1995. Description of *Electrogena calabra* n. sp., a new species from Southern Italy (Ephemeroptera, Heptageniidae). – Annales de Limnologie **31**(1): 29−34. doi:10.1051/limn/1995001

Belfiore C. 1996. Identification and discrimination of *Electrogena* species by numerical methods (Ephemeroptera: Heptageniidae). – Systematic Entomology **21**(1): 1−13. doi:10.1111/j.1365-3113.1996.tb00595.x

Belfiore C. 1997. Taxonomic characters and discrimination of species in the genus *Electrogena* Zurwerra & Tomka (Ephemero-

ptera, Heptageniidae). Pp. 427–433 in: LANDOLT P., SARTORI M. (eds), Ephemeroptera & Plecoptera: Biology, Ecology, Systematics. – MTL, Fribourg.

BELFIORE C., DESIO F. 1995. Taxonomy and distribution of *Electrogena ujhelyii* (SOWA, 1981). – Annalen des Naturhistorischen Museums in Wien, Serie B für Botanik und Zoologie **97**(B): 151–154.

BELFIORE C., SCILLITANI G., PICARIELLO O., CATAUDO A. 1997. Morphological and electrophoretic evidence for a new species of *Electrogena* from central Italy: Description of *E. lunaris* sp. n. (Ephemeroptera: Heptageniidae). – Aquatic Insects **19**(3): 129–140. doi:10.1080/01650429709361647

BELFIORE C., HAYBACH A., KLONOWSKA-OLEJNIK M. 1999. Taxonomy and phenetic relationships of *Electrogena affinis* (Eaton, 1883) (Ephemeroptera : Heptageniidae). – Annales de Limnologie - International Journal of Limnology **35**(4): 245–256. doi:10.1051/limn/1999033

BELFIORE C., SARTORI M. 1999. Redescription of *Electrogena galileae* (Demoulin, 1973) (Ephemeroptera, Heptageniidae). – Revue Suisse de Zoologie **106**(4): 1025–1034. doi:10.5962/bhl.part.80114

BELFIORE C., TANATMIS M., KAZANCI N. 2000. Taxonomy of *Electrogena antalyensis* (Kazanci & Braasch, 1986) (Ephemeroptera, Heptageniidae). – Aquatic Insects **22**(4): 261–270. doi:10.1076/0165-0424(200010)22:4;1-Y;FT261

BONHOMME V., PICQ S., GAUCHEREL C., CLAUDE J. 2014. Momocs: Outline Analysis Using R. – Journal of Statistical Software **56**(13): 1–24. doi:10.18637/jss.v056.i13

BOUCKAERT R., HELED J., KÜHNERT D., VAUGHAN T., WU C.-H., XIE D., SUCHARD M.A., RAMBAUT A., DRUMMOND A.J. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. – PLoS Computational Biology **10**(4): e1003537. doi:10.1371/journal.pcbi.1003537

BRITTAIN J.E. 2008. Mayflies, biodiversity and climate change. Pp. 1–14 in: HAUER F.R., STANFORD J., NEWELL R. (eds), International Advances in the Ecology, Zoogeography and Systematics of Mayflies and Stoneflies. – University of California Publications in Entomology. doi:10.1525/california/9780520098688.003.0001

CARDONI S., TENCHINI R., FICULLE I., PIREDDA R., SIMEONE M.C., BELFIORE C. 2015. DNA barcode assessment of Mediterranean mayflies (Ephemeroptera), benchmark data for a regional reference library for rapid biomonitoring of freshwaters. – Biochemical Systematics and Ecology **62**: 36–50. doi:10.1016/j.bse.2015.07.035

CHESTERS D. 2013. collapsetypes_v4.5. – URL <https://sourceforge.net/projects/collapsetypes> [accessed 01 December 2017].

CLAUDE J. 2008. Morphometrics with R. 1st edn. – Springer-Verlag, New York. 317 pp. doi:10.1007/978-0-387-77789-4

DARRIBA D., TABOADA G.L., DOALLO R., POSADA D. 2012. jModelTest 2: more models, new heuristics and parallel computing. – Nature Methods **9**(8): 772. doi:10.1038/nmeth.2109

DUDGEON D., ARTHINGTON A.H., GESSNER M.O., KAWABATA Z.-I., KNOWLER D.J., LÉVÊQUE C., NAIMAN R.J., PRIEUR-RICHARD A.-H., SOTO D., STIASSNY M.L.J., SULLIVAN C.A. 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. – Biological Reviews **81**(2): 163–182. doi:10.1017/S1464793105006950

FAVRET C., SIERACKI J.M. 2016. Machine vision automated species identification scaled towards production levels. – Systematic Entomology **41**(1): 133–143. doi:10.1111/syen.12146

FOLMER O., BLACK M., HOEH W., LUTZ R., VRIJENHOEK R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. – Molecular Marine Biology and Biotechnology **3**(5): 294–299.

FUJISAWA T., BARRACLOUGH T.G. 2013. Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent approach: a revised method and evaluation on simulated data sets. – Systematic Biology **62**(5): 707–724. doi:10.1093/sysbio/syt033

GATTOLLIAT J.-L., CAVALLO E., VUATAZ L., SARTORI M. 2015. DNA barcoding of Corsican mayflies (Ephemeroptera) with implica-

tions on biogeography, systematics and biodiversity. – Arthropod Systematics & Phylogeny **73**(1): 3–18.

GODUNKO R.J. 2000. Little known species of the genera *Rhithrogena* and *Electrogena* (Ephemeroptera, Heptageniidae) from Ukraine. – Vestnik Zoologii **14**: 60–66.

GODUNKO R.J., PROKOPOV G.A., KLONOWSKA-OLEJNIK M. 2002. Complementary description of the winged stages of *Electrogena braaschi* (Ephemeroptera, Heptageniidae). – Vestnik Zoologii **36**(5): 73–76.

GUINDON S., GASCUEL O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. – Systematic Biology **52**(5): 696–704. doi:10.1080/10635150390235520

HAYBACH A. 1999. Beitrag zur Larvaltaxonomie der *Ecdyonurus venosus*-Gruppe in Deutschland. – Lauterbornia **37**: 113–150.

HAYBACH A. 2008. Egg structure morphology gives evidence for a cryptic subspecies of *Electrogena lateralis* (Curtis, 1834) in Southern Europe: *Electrogena lateralis concii* (GRANDI, 1953) nov. stat. [Ephemeroptera, Heptageniidae]. – Ephemera **9**(1): 1–7.

KATOH K., STANDLEY D.M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. – Molecular Biology and Evolution **30**(4): 772–780. doi:10.1093/molbev/mst010

KŁONOWSKA-OLEJNIK M. 2004. Redescription of *Electrogena quadrilineata* (Landa, 1969) from type material (Ephemeroptera, Heptageniidae). – Aquatic Insects **26**(2): 85–95. doi:10.1080/01650420412331325828

KLUGE N. 2004. The phylogenetic system of Ephemeroptera. 1st edn. – Kluwer Academic Publishers, Dordrecht. 442 pp. doi:10.1007/978-94-007-0872-3

LANDA V. 1969. Fauna ČSSR: Jepice - Ephemeroptera. – Academia, Praha. 352 pp.

LANDA V., SOLDÁN T. 1982. *Ecdyonurus samalorum* sp. n. from Czechoslovakia (Ephemeroptera, Heptageniidae). – Acta Entomologica Bohemoslovaca **79**: 31–36.

LARIOS N., DENG H., ZHANG W., SARPOLA M., YUEN J., PAASCH R., MOLDENKE A., LYTLE D.A., CORREA S.R., MORTENSEN E.N., SHAPIRO L.G., DIETTERICH T.G. 2008. Automated insect identification through concatenated histograms of local appearance features: feature vector generation and region detection for deformable objects. – Machine Vision and Applications **19**(2): 105–123. doi:10.1007/s00138-007-0086-y

MONAGHAN M.T., WILD R., ELLIOT M., FUJISAWA T., BALKE M., INWARD D.J.G., LEES D.C., RANAIVOSOLO R., EGGLETON P., BARRACLOUGH T.G. 2009. Accelerated species inventory on Madagascar using coalescent-based models of species delineation. – Systematic Biology **58**(3): 298–311. doi:10.1093/sysbio/syp027

RAMBAUT A., DRUMMOND A.J. 2015. LogCombiner v2.3. – URL: <http://beast.community/logcombiner> [accessed 01 December 2017].

RAMBAUT A., DRUMMOND A.J., XIE D., BAELE G., SUCHARD M.A. 2018. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. – Systematic Biology **67**(5): 901–904. doi: 10.1093/sysbio/syy032

RIGHETTI B., BELFIORE C. 2000. Additions à la faune des Éphémères de France (5): *Electrogena grandiae* (Belfiore, 1981) et comparaison biométrique de différentes populations [Ephemeroptera, Heptageniidae]. – Ephemera **1**(2): 123–130.

RUTSCHMANN S., GATTOLLIAT J.-L., HUGHES S.J., BÁEZ M., SARTORI M., MONAGHAN M.T. 2014. Evolution and island endemism of morphologically cryptic *Baetis* and *Cloeon* species (Ephemeroptera, Baetidae) on the Canary Islands and Madeira. – Freshwater Biology **59**(12): 2516–2527. doi:10.1111/fwb.12450

RUTSCHMANN S., DETERING H., SIMON S., FREDSLUND J., MONAGHAN M.T. 2017a. DiscoMark: Nuclear marker discovery from orthologous sequences using draft genome data. – Molecular Ecology Resources **17**: 257–266. doi: 10.1111/1755-0998.12576

RUTSCHMANN S., DETERING H., SIMON S., FUNK D.H., GATTOLLIAT J.-L., HUGHES S.J., RAPOSEIRO P.M., DESALLE R., SARTORI M., MONAGHAN M.T. 2017b. Colonization and diversification of aquatic insects on three Macaronesian archipelagos using 59 nuclear loci

derived from a draft genome. – Molecular Phylogenetics and Evolution **107**: 27–38. doi:10.1016/j.ympev.2016.10.007

SCHOENEMUND E. 1930. Eintagsfliegen oder Ephemeroptera. 19th edn. – Fischer Verlag, Jena. 106 pp.

SHAPIRO B., RAMBAUT A., DRUMMOND A.J. 2006. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. – Molecular Biology and Evolution **23**(1): 7–9. doi:10.1093/molbev/msj021

SROKA P., GODUNKO R.J. 2012. Mayflies of the Caucasus Mountains. I. A new species of the genus *Electrogena* Zurwerra & Tomka, 1985. – Zootaxa **3222**: 28–45. doi:10.11646/zootaxa.4231.1.4

SOWA R. 1981. Taxonomy and ecology of *Ecdyonurus ujhelyii* sp. n. (Ephemeroptera, Heptageniidae) from the tributaries of Lake Balaton. – Acta Hydrobiologica **23**(4): 375–380.

TAMURA K., PETERSON D., PETERSON N., STECHER G., NEI M., KUMAR S. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. – Molecular Biology and Evolution **28**(10): 2731–2739. doi:10.1093/molbev/msr121

VUATAZ L., SARTORI M., WAGNER A., MONAGHAN M.T. 2011. Toward a DNA taxonomy of Alpine *Rhithrogena* (Ephemeroptera: Heptageniidae) using a Mixed Yule-Coalescent Analysis of mitochondrial and nuclear DNA. – PLoS ONE **6**(5): e19728. doi:10.1371/journal.pone.0019728

VUATAZ L., RUTSCHMANN S., MONAGHAN M.T., SARTORI M. 2016. Molecular phylogeny and timing of diversification in Alpine *Rhithrogena* (Ephemeroptera: Heptageniidae). – BMC Evolutionary Biology **16**(1): 194. doi:10.1186/s12862-016-0758-1

WAGNER A., VUATAZ L., SARTORI M. 2017. *Electrogena brulini* sp. nov. and *E. vipavensis* Zurwerra & Tomka, 1986 syn. nov. revealed by integrative taxonomy of *E. gridellii* (Grandi, 1953) (Ephemeroptera: Heptageniidae). – Zootaxa **4362**(3): 359. doi: 10.11646/zootaxa.4362.3.3

WAKIMURA K., TAKEMON Y., TAKAYANAGI A., ISHIWATA S., WATANABE K., TANIDA K., SHIMIZU N., KATO M. 2016. Characterization of genes for histone H3, 18S rRNA, and cytochrome oxidase subunit I of East Asian mayflies (Ephemeroptera). – DNA Barcodes **4**(1): 1–25. doi:10.1515/dna-2016-0001

WEN C., GUYER D. 2012. Image-based orchard insect automated identification and classification method. – Computers and Electronics in Agriculture **89**: 110–115. doi:10.1016/j.compag.2012.08.008

YANAI Z., SARTORI M., DOR R., DORCHIN N. 2017. Molecular phylogeny and morphological analysis resolve a long-standing controversy over generic concepts in Ecdyonurinae mayflies (Ephemeroptera: Heptageniidae). – Systematic Entomology **42**(1): 182–193. doi:10.1111/syen.12203

ZURWERRA A., TOMKA I. 1985. *Electrogena* gen. nov., eine neue Gattung der Heptageniidae [Ephemeroptera]. – Entomologische Berichte Luzern **13**: 99–104.

ZURWERRA A., TOMKA I. 1986. Drei neue Arten der Gattung *Electrogena* Zurwerra & Tomka 1985, aus Südeuropa (Ephemeroptera, Heptageniidae). – Bulletin de la Societé Fribourgeoise des Sciences Naturelles **75**(1/2): 216–230.

ZURWERRA A., METZLER M., TOMKA I. 1987. Biochemical systematics and evolution of the European Heptageniidae (Ephemeroptera). – Archiv für Hydrobiologie **109**(4): 481–510.

7. Appendix

7.1. Quantitative (numeric) characters

N_LBB – Number of median bristles on dorsal side of labrum. Only the short and stout bristles, maximally 2.5 × longer than first, medial bristles, were counted. These usually reach to the frontal edge of labrum – the row then continues up to the lateral lobes of the labrum, surrounded by other thin and long bristles. This character was used for the first time to describe the *Electrogena* larvae.

N_CBS – Number of comb-shaped bristles on fore margin of galea-lacinia. If possible, the mean of the number of comb-shaped bristles on the left and right galea-lacinia was calculated. For details see BELFIORE (1996) and HAYBACH (1999).

N_TCB – Number of pointed teeth on 5th comb-shaped bristle from medially. If possible, the mean of the number of pointed teeth on the 5th comb-shaped bristle on the left and right galea-lacinia was calculated. For details see BELFIORE (1996) and HAYBACH (1999).

N_OUT – Number of hairs near outer edge of galea-lacinia, from palpus to fore margin. Character used in BELFIORE (1996), arrangement of hairs in the *Electrogena* species slightly differs from arrangement of hairs in genus *Ecdyonurus* pictured in HAYBACH (1999). Entire lateral edge of galea-lacinia taken into account, no different types of hairs (in HAYBACH (1999) defined as N_DOR and N_OUT) observed.

N_PLP – Number of hairs near fore margin of first segment of maxillary palp. Arrangement of hairs is similar to HAYBACH (1999), the

character also used in BELFIORE (1996). If possible, average number of hairs in left and right maxillary palp calculated.

N_PLS – Number of stout bristles on hind margin of first segment of maxillary palp. Arrangement of bristles is similar to HAYBACH (1999). If possible, average number of bristles in left and right maxillary palp calculated. This character was used for the first time to describe the *Electrogena* larvae.

N_PLH – Number of long hairs on hind margin of first segment of maxillary palp. Arrangement of hairs is similar to HAYBACH (1999). Only long and thin hairs near the basis of first segment of maxillary palp taken into account. If possible, average number of hairs in left and right maxillary palp calculated.

N_PLBas – Number of hairs at base of maxillary palp. Arrangement of hairs is similar to HAYBACH (1999). If possible, average number of hairs in left and right maxilla calculated.

N_VEN – Number of hairs on ventral basal part of maxilla. Arrangement of hairs is similar to HAYBACH (1999). If possible, average number of hairs in left and right maxilla calculated. This character was used for the first time to describe the *Electrogena* larvae.

N_MPB(l) – Number of bristles near left prostheca. Stout, long bristles following incisive parts counted. This character was used for the first time to describe the *Electrogena* larvae.

N_MPB(r) – Number of bristles near right prostheca. Stout, long bristles following incisive parts counted.

N_LPH_prox – Number of bristles on hind margin of first segment of labial palp – proximal part. Arrangement of bristles slightly differs from arrangement stated in HAYBACH (1999). Character N_LPH in this article describes number of bristles and hairs in dorsolateral

part of labial palp, in the article counted for bristles and hairs separately. In our approach, only one type of bristles was recognised and counted. Area covered by bristles in Central European *Electrogena* species is divided into two discontinued parts – proximal and distal part. Number of bristles of each part counted separately. This character was used for the first time to describe the *Electrogena* larvae.

N_LPH_dist – Number of bristles on hind margin of first segment of labial palp - distal part. Details discussed above. This character was used for the first time to describe the *Electrogena* larvae.

N_CLW – Number of teeth on pretarsal claws. If possible, average of fore, mid and hind leg counted.

N_BVF – Number of bristles on ventral side of femora near hind margin. If possible, number of these on profemora counted.

N_HFF – Number of long hairs (at least twice as long as neighbouring bristles) on fore margin of femora. If possible, number of these on profemora counted. This character was used for the first time to describe the *Electrogena* larvae.

7.2. Quantitative (ratio) characters

R_LBR – Total width of labrum divided by mean width of lateral lobes of labrum. For details see BELFIORE (1996).

R_GLA – Distance between glossae (outer distance divided by inner distance between glossae). For details see BELFIORE (1996).

R_GLB – Width of glossae (outer distance divided by width of glossae). For details see BELFIORE (1996).

R_1GI – Relative width of 1st gill plate.

R_4GI – Relative width of 4th gill plate.

R_7GI – Relative width of 7th gill plate.

R_3FEM – Relative width of metafemora (length/width ratio of metafemora). This character was used for the first time to describe the *Electrogena* larvae.

7.3. Qualitative characters

S_BFE(i) – Shape of setae on upper surface of profemora. Four types of this character were recognised: **pointed**, **round pointed**, **bottle-like**, and **spatula-like**. Shape of setae evaluated in central area of upper side of profemora. In case of indecisive type of character, both types (states of the character) were recorded.

S_BFE(iii) – Shape of setae on upper surface of metafemora. For details see above (S_BFE(i)).

S_MFE – Shape of metafemora. Two types of this character recognised: **symmetrical** and **asymmetrical**. State of this character estimated by expert judgement; evaluation was based on comparing convexity of fore and hind margin of metafemora. This character was used for the first time to describe the *Electrogena* larvae.

S_TAR – Marking on the tarsi. Four types of this character were recognised: **light**, **unicolour dark**, **distally darkened** and **distally and proximally darkened**.

S_HEM – Markings on the head. Three types were recognized: **unicolor brown head** with no light spots on front edge of head capsule (type A); head with **two lighter smudges** in the middle on frontal edge of head capsule (type B); head with distinct **light spots** on frontal edge of head capsule (type C). Overall colour of head (from yellowish to dark brown) and light dots or smudges in central area of head capsule not taken into account.

S_HLB – Extension of pilosity on lateral lobes. Two types were recognised: **pilosity of hypopharynx does not overlap apex** of superlinguae, or **pilosity of hypopharynx overlaps apex** of superlinguae. Only long, fine setae taken into account.

Authors' contributions

M.P., M.S. and M.N. contributed to the material sampling, M.P. processed and sequenced all material, S.R. performed molecular part of analyses, M.P. and R.J.G. collaborated on morphological part of analyses, all authors collaborated on manuscript preparation with M.P. as a leading author and S.Z. as a supervisor.

Electronic Supplement File

at http://www.senckenberg.de/arthropod-systematics

File 1: polasek&al-electrogenaheptageniidae-asp2018-electronic supplement-1.pdf — **Table S1.** A list of all analysed specimens with country and locality specification and GenBank accession numbers for cox1 mitochondrial gene sequences. — **Table S2.** Numerical and qualitative characters of European *Electrogena* species based on available publications. — **Table S3.** Minimum, mean and maximum for all numerical characters used in this study. — **Table S4.** Percentage representation of qualitative character states used in this study. — **Fig. S1.** Boxplots of discrete numerical characters for defined GMYC OTUs. — **Fig. S2.** Boxplots of ratio numerical characters for defined GMYC OTUs. — **Fig. S3.** The histograms of percental representation of qualitative character states.

GenBank accession numbers

at https://www.ncbi.nlm.nih.gov

GenBank accession numbers for all used sequences are available in Supplement file. Continuous series of 69 GenBank accession numbers start with MH093830 and ends with MH093898.