First molecular phylogeny of *Agrilus* (Coleoptera: Buprestidae), the largest genus on Earth, with DNA barcode database for forestry pest diagnostics

I. Kelnarova[1], E. Jendek[2], V.V. Grebennikov[3]* and L. Bocak[1]

[1]Department of Zoology, Faculty of Science UP, Olomouc, Czech Republic:
[2]Department of Forest Protection and Entomology, Faculty of Forestry and Wood Sciences, Czech University of Life Sciences, Kamýcká 1176, CZ–165 21, Prague 6–Suchdol, Czech Republic: [3]Canadian Food Inspection Agency, 960 Carling Avenue, Ottawa, ON K1A 0Y9, Canada

Abstract

All more than 3000 species of *Agrilus* beetles are phytophagous and some cause economically significant damage to trees and shrubs. Facilitated by international trade, *Agrilus* species regularly invade new countries and continents. This necessitates a rapid identification of *Agrilus* species, as the first step for subsequent protective measures. This study provides the first DNA reference library for ∼100 *Agrilus* species from the Northern Hemisphere based on three mitochondrial markers: *cox1*–5′ (DNA barcode fragment), *cox1*–3′, and *rrnL*. All 329 *Agrilus* records available in the Barcode of Life Database format, including specimen images and geo data, are released through a public dataset 'Agrilus1 329' available at: dx.doi.org/10.5883/DS-AGRILUS1. All *Agrilus* species were identified using adult morphology and by using molecular phylogenetic trees, as well as distance- and tree-based algorithms. Most DNA-based species limits agree well with the morphology-based identification. Our results include cases of high intraspecific variability and multiple species para- and polyphyly. DNA barcoding is a powerful species identification tool in *Agrilus*, although it frequently fails to recover morphologically-delimited *Agrilus* species-group. Even though the current three-gene database covers only ∼3% of the known *Agrilus* diversity, it contains representatives of all principal lineages from the Northern Hemisphere and represents the most extensive dataset built for DNA-delimited species identification within this genus so far. Molecular data analyses can rapidly and cost-effectively identify an unknown sample, including immature stages and/or non-native taxa, or species not yet formally named.

Keywords: invasive species, jewel beetles, mtDNA, phylogeny, species delimitation

Introduction

The cosmopolitan jewel beetle genus *Agrilus* Curtis, with its 3213 valid nominal species, is the largest conventionally recognized genus of the Animal Kingdom (as of July 31 2017; Bellamy, 2008; Zoological Record®, Thomson Reuters). All *Agrilus* are strictly phytophagous and their astonishing diversity (figs 2a–i and 3a–i) might perhaps be linked to numerous shifts among their host plants, which is an important diversification factor in many phytophagous insects (Drès & Mallet, 2002; Nosil & Mooers, 2005; Egan *et al.*, 2008). While *Agrilus* adults feed on leaves, their larvae (fig. 1c) develop on living subcortical tissues of trees and shrubs to an extent sufficient to kill a host, especially when it has been already

*Author for correspondence
Phone: +1-613-759-7519
Fax: +1-613-759-6938
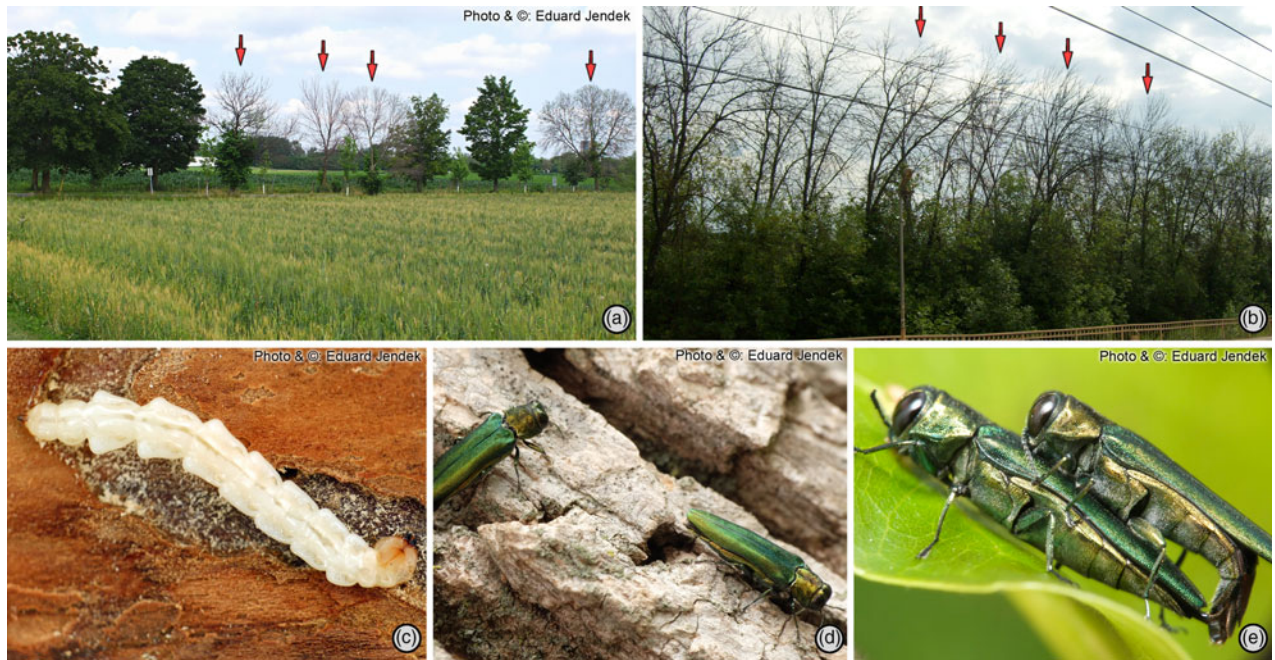E-mail: vasily.grebennikov@inspection.gc.ca

Fig. 1. *Agrilus planipennis* (=Emerald Ash Borer), the most costly insect forestry pest in human history. (a) and (b): examples of damage: (a) North America (Canada, Ottawa); (b) Europe (Russia, Moscow Region); larva (c) and adults (d) and (e).

weakened by abiotic factors (Wargo, 1977; Dunn *et al.*, 1986; Jendek & Poláková, 2014). This economically significant effect is achieved by inducing subcortical necrosis that disrupts water and nutrient transport (Vansteenkiste *et al.*, 2004). Moreover, *Agrilus* species have a proven invasive potential facilitated by their long-living immature stages easily transported with wood, while the adults are capable of long active flights. As a consequence, some *Agrilus* species became important invasive alien pests (Haack *et al.*, 1983; Akers *et al.*, 1986; Gordon *et al.*, 1997; Everett, 2000) after being accidentally introduced to an area free of natural enemies (Dunn *et al.*, 1986; Jones *et al.*, 1993; Gibbs & Greig, 1997; Aukema *et al.*, 2011). Consequently, *Agrilus* beetles constitute a high-risk group of organisms for quarantine entomologists, particularly in North America, and on the level similar to those of longhorned beetles (Cerambycidae) and bark and ambrosia beetles (Curculionidae: Scolytinae and Platypodinae).

Indeed, human-mediated range expansions are well documented in *Agrilus* with at least 11 Eurasian species having been introduced to North America: *Agrilus cuprescens* (Ménétriés), *Agrilus cyanescens* (Ratzeburg), *Agrilus derasofasciatus* Lacordaire, *Agrilus hyperici* (Creutzer), *Agrilus pilosovittatus* Saunders, *Agrilus planipennis* Fairmaire, *Agrilus ribesi* Schaefer, *Agrilus sinuatus* (Olivier), *Agrilus smaragdifrons* Ganglbauer, *Agrilus subrobustus* Saunders, and *Agrilus sulcicollis* Lacordaire (Jendek & Grebennikov, 2009; Jendek, 2016; Hoebeke *et al.*, 2017). The European species, *Agrilus angustulus* (Illiger), might have been introduced to the Russian Far East, while the North American species, *Agrilus bilineatus* (Weber), is the first among the Nearctic species to reach the Palaearctic Region (based on the single record in Turkey, Jendek 2016). Among these adventive *Agrilus* taxa, one has become most popular and feared. This is the infamous *A. planipennis*, the Emerald Ash Borer (figs 1a–e), which is a prime example of

a large-scale and continent-wide biological and economic disaster caused by an introduction of a 'bad' jewel beetle to two new continents (North America and Europe; fig. 1a, b, respectively). Damage caused by this notorious pest was partly ascribed to the high susceptibility of at least the North American Ash trees to beetle's attacks, and partly due to the absence of natural enemies (Haack *et al.*, 2002; Liu *et al.*, 2003). The negative economic consequences caused by this beetle in North America have been indeed enormous and variously estimated to exceed 10 billion USD (Kovacs *et al.*, 2011; Herms & McCullough, 2014).

Rapid detection and taxonomic identification of an invasive species is the foremost step that leads to appropriate control measures to follow. The lack of *Agrilus* taxonomic expertise and absence of adequate diagnostic tools result in a significant delay between the establishment of an invasive species and its first positive identification (Jendek & Grebennikov, 2011). For example, it took more than 5 years to detect and identify *A. planipennis* in North America (Haack *et al.*, 2002). Still worst, the occurrence of the introduced *A. ribesi* had been overlooked in the USA and Canada for almost a century, while the repeatedly observed damage was erroneously assigned to a native species (Jendek *et al.*, 2015). Since the number of non-native pests raises continuously (Aukema *et al.*, 2010), it becomes of prime economic importance to develop DNA-based pest diagnostic tools to facilitate their rapid and reliable identification.

The sheer number of species within the enormously diverse genus *Agrilus* generates its unique challenge, which in turn inhibits an efficient use of generic taxonomy. The current internal classification of the genus is highly inconsistent and incomplete. All 35 nominal *Agrilus* subgenera proposed so far (e.g., Alexeev, 1998; Bellamy, 2008) were introduced on a regional (mainly European) and non-phylogenetic basis by

Fig. 2. *Agrilus*, diversity of Northern Hemisphere species. (a) *A. acutipennis* (Canada, Ontario); (b) *A. anxius* (Canada, Ontario); (c) *A. cyaneoniger* (Russia, Primorsky Krai); (d) *A. dureli* (China, Beijing); (e) *A. egeniformis* (Canada, Ontario); (f) *A. fleischeri* (China, Jilin); (g) *A. granulatus* (Canada, Ontario); (h) *A. guerini* (Slovakia); (i) *A. hyperici* (Slovakia).

using the vague criterion of general similarity among adult beetles. In the absence of a comprehensive worldwide taxonomic treatment of the genus, the vast majority of nominal *Agrilus* species remain unassigned to any of these 'subgenera'. As a result, the introduction of formal subgenera arguably caused more inconsistencies and taxonomic confusion, compared with not being introduced at all (Jendek & Grebennikov, 2011). The undeveloped and rigid (in terms of zoological nomenclature) sub-generic approach has been recently substituted by the use of informal species-groups serving as temporary taxonomic labels preventing the proliferation of unnecessary genus-group names (Jendek & Grebennikov, 2011).

The species-level identification of *Agrilus* beetles outside of a few well-studied regions, such as Europe and temperate North America, is often a significant challenge and in many cases is nearly, or outright, impossible. This is due to the immense number of *Agrilus* species, as well as to the much-constrained adult morphological variability resulting in a high number of externally similar species (Jendek & Grebennikov, 2011). As a result, reliable *Agrilus* identification requires expert skills and an extensive reference collection. An experienced entomologist not particularly specialized in jewel

beetles is normally able to identify adults of locally native *Agrilus* species, but will almost always fail to do so for any larva or for non-native adults, particularly from a geographically distant region. Furthermore, the presence of a well-curated and extensive voucher collection of *Agrilus* will always be required for such identifications. Overall, putting a correct species name on an *Agrilus* beetle, adult or immature, is a task beyond the current diagnostic capacity of many entomological professionals, research organizations or even that of some developed countries (Jendek & Grebennikov, 2011).

Recently mitochondrial DNA-based species identification methods have become increasingly important as a practical alternative to the classical morphology-based identification (e.g., Hebert *et al.*, 2003; Riedel *et al.*, 2013a, b; Ashfaq & Hebert, 2016). Although there is a possibility that nuclear-encoded mitochondrial pseudogenes can be amplified instead of the mitochondrial target fragment or that several sequence types might be present in cells of a single individual, the method can be used when high numbers of individuals and representatives of different populations are sequenced to identify the dominant protein coding haplotype (Kang *et al.*, 2016). The sequencing of short mtDNA fragments with conservative

Fig. 3. *Agrilus*, diversity of Northern Hemisphere species. (a) *A. obsoletoguttatus* (Canada, Ontario); (b): *A. pensus* (Canada, Ontario); (c) *A. politus* (Canada, Ontario); (d) *A. pseudocoryli* (Canada, Ontario); (e) *A. ribbei* (Russia, Primorsky Krai); (f) *A. ruficollis* (Canada, Ontario); (g) *A. smaragdinus* (China, Jilin); (h) *A. viridis* (Slovakia); (i) *A. vittaticollis* (Canada, Ontario).

primers has been repeatedly demonstrated to be a rapid, reliable and cost-efficient alternative, even though a subsequent critical evaluation of the results by using classical morphology by a taxonomic expert remains desirable. Besides being rapid and expertise-free, DNA-based methods enable the identification of immature stages, which may be the only available representatives of a species when the economically significant damage is identified for the first time (Ahrens *et al.*, 2007). In ecological and biodiversity research, putative biological species can be preliminarily delimited and identified using various DNA-based methods. The simplest of such methods is the one based on pairwise genetic distances. The species-delimitation results, however, might differ substantially when various thresholds are applied (Hebert *et al.*, 2003). Since the method was proposed, 3% of the uncorrected pairwise distance threshold has been repeatedly rejected as a 'universal' value, mainly because extreme deviations from this rate have been detected in either direction. Furthermore, the *cox1*–5′ fragment (=DNA barcode) of morphologically or ecologically distinct species might be identical or, conversely, genetically highly divergent populations might be detected within the same 'good' morphological species (Baselga *et al.*,

2013; Zahiri *et al.*, 2014; Li *et al.*, 2015; Kusy *et al.*, 2018). Since such information for *Agrilus* beetles has been limited, in this study we attempt to present a large database of mtDNA fragments to evaluate the levels of genetic variability within this genus.

The tree-based methods of species identification (Satler *et al.*, 2013) form a phylogenetically-sound alternative to the threshold-based approach. They integrate morphological and molecular methods in insect systematics of predominantly highly diversified clades, with the aim of producing robust species-level phylogenetic trees (e.g., Pons *et al.*, 2006; Riedel *et al.*, 2013b; Bocek & Bocak, 2016). Unlike the DNA-barcode threshold approach, these methods were not primarily designated for the routine identification of an unknown sample against the reference database. Instead, they enable the positioning of an unknown sample with a high degree of confidence in the phylogenetic framework and among its closest relatives (Vogler & Monaghan, 2007; Monaghan *et al.*, 2009). So performed, the tree-based methods are superior in their diagnostic capacity to the distance-based matching approach, partly because they can handle unknown unknowns, that are representatives of yet undescribed species. Instead of a

numerical and somewhat arbitrary genetic distance threshold, the tree-based methods use the shape of the phylogenetic tree, which might be either a dated tree (obtained with the General Mixed Coalescent model, GMYC; Pons *et al.*, 2006; Fujisawa & Barraclough, 2013) or a phylogram (obtained with the Bayesian Poisson Tree process, Zhang *et al.*, 2013). The results of all tree-based analyses are not necessarily identical when different methods are used (Carstens *et al.*, 2013), and various levels of success have been reported in unrelated animal lineages (Baselga *et al.*, 2013; Li *et al.*, 2015; Bocek & Bocak, 2016). Nevertheless, the performance of the tree-based methods is at least comparable with the identification by a trained non-specialist, while the even greater reliability of the tree-based identification can be expected with a large reference dataset. Another important advantage of a tree-based approach is the possibility of identifying the geographic origin of a non-native sample, either within the intraspecific variability (if samples from different parts of the range are available) or, when a lineage has a phylogeographic structure, the closest relatives can be expected in the same region. In the present work, therefore, we will attempt to investigate the applicability of the tree-based identification for *Agrilus* pests by recovering geographically-delimited clades and morphologically predefined groups of closely related species.

In summary, the principal aim of the present study is to report and make publicly available the extensive database of three mitochondrial DNA fragments and new barcode sequences for a large number of *Agrilus* species analyzed so far. The recovered phylogenetic trees provide information on relationships among principal *Agrilus* lineages and the newly generated data can be used for DNA-based identification of unknown *Agrilus* samples originating from anywhere within the temperate zone of the Northern Hemisphere and irrespective of the beetle's life stage. Furthermore, we test the performance of various DNA-based species delimitation methods against the dataset of morphologically identified *Agrilus* species. Specifically, we test if the results are similar when different species delimitation methods are applied and, additionally, which particular species is conflictingly delimited using morphology vs DNA data. If a single species is represented by multiple distantly related lineages, the current species concept is unacceptable in the biological classification (Vences *et al.*, 2013) and the limits of such a species need further study. Additionally, we investigate the extent in which the closely related species might occur in distant zoogeographical regions, that is in Europe, in East Asia and in Northern America. Overall this work should serve as a starting point for building the densely sampled public online *Agrilus* database in support of the DNA-based identification of economically significant pests from the temperate zone of the Northern Hemisphere.

Material and methods

Taxon sampling

In the course of this study a total of 475 *Agrilus* individuals of over 100 species were collected by various collectors in 2000–2015 in Eastern Europe and in Central Europe north of the Alps (111 individuals), the Mediterranean Region (57 individuals), Russian Far East (111 individuals), China, Korea and Japan (23 individuals), the Oriental Region (11 individuals) and North America (154 individuals). All these specimens were identified to species level by one of us (EJ). Voucher numbers, NCBI GenBank accession numbers and detailed locality and host plant data are given in table S1. The voucher specimen numbers are used to identify any sequenced individual in the published trees herein and in the GenBank database. All 329 *Agrilus* records available in the Barcode of Life Database (=BOLD) format, including specimen images (fig. S3) and geo data, are released through a public dataset 'Agrilus1 329' available at: dx.doi.org/10.5883/DS-AGRILUS1. These specimens have voucher numbers (column 2 in table S1 in format CNC0000) having the same last four digits, as their BOLD Sample ID (in format CNCCOLVG00000000). The dry-mounted sequenced specimens are deposited in the voucher collections located at the authors' home institutions. The outgroup was composed of the most likely closely related non-*Agrilus* Agrilini jewel beetles from the genera *Coraebus* Gory & Laporte, *Meliboeus* Deyrolle, *Nalanda* Thery, and *Trachys* Fabricius.

DNA isolation, amplification and sequencing

All sequenced specimens were killed and preserved in 96% ethanol and stored at −20 °C until DNA extraction. The samples designated by voucher numbers beginning with CNC were sequenced for the DNA barcoding fragment in the Biodiversity Institute of Ontario, Canada following the standard DNA barcoding protocol (Ivanova *et al.*, 2006; Grebennikov *et al.*, 2017). Additional sequencing of the same specimens for two other DNA markers, as well as sequencing of all samples designated by voucher numbers beginning with "EJ" or "A" was performed in the Laboratory of Molecular Systematics, Palacky University, Olomouc, Czech Republic. For these, genomic DNA was extracted from metathoracic muscles using either the phenol chloroform method or the Qiagen DNeasy® Blood & Tissue Kit (Qiagen Inc.). Extraction yield was measured using a NanoDrop-1000 Spectrophotometer. Three fragments of mitochondrial DNA (mtDNA) were amplified: *rrnL* fragment using primers 16a and 16b and alternatively the primer ND1A, which is located in the *nad1* mtDNA gene; the *cox1*–3′ fragment using primers Pat and Jerry/JerM, and the *cox1*–5′ (=DNA barcode) mtDNA fragment using primers LCO1490 K and HCO2198 K (Simon *et al.*, 1994; Bocak *et al.*, 2008). Polymerase chain reaction (PCR) settings and cycle sequencing conditions were those reported by Bocak *et al.* (2008), whereas the PCR products were purified using PCRu96 Plates (Millipore Inc.) and sequenced by an ABI 3130 automated sequencer using the Big Dye Sequencing Kit 1.1. Altogether, 780 new sequences were produced for this study: 166 sequences for *rrnL* mtDNA, 170 for *cox1*–3′ mtDNA, and 444 for *cox1*–5′ (DNA barcode fragment) mtDNA (table S1).

Sequence editing and alignment

Sequences were edited using Sequencher 4.10.1 (Gene Codes Corp.). The length conservative protein coding sequences were aligned using ClustalX 2.1 (Thompson *et al.*, 1997) under default parameter settings and checked for amino acid reading frames to identify and exclude potential pseudogenes from further analyses. The studied group represents a single genus of relatively closely related species and when the nucleotide sequences were translated in amino acid sequence, we identified only a low variability. The pseudogene would be easily identified by the highly divergent amino acid sequence. If any pseudogene is present in the current Agrilus dataset, its mutations have to be synonymous with the amino acid translation and the length of the sequence

could not change. Such pseudogene cannot be detected by available methods, but on the other hand, the phylogenetic noise potentially introduced by its presence is negligible.

The length variable rrnL fragment was aligned using Mafft 7 (Katoh & Standley, 2013) under default parameter settings. The rrnL mtDNA alignment consisted of 883 aligned positions. As all ingroup taxa belong to a single genus, they are highly similar and the maximum number of contiguous non-conserved positions reached eight nucleotides. As such short length variable blocks do not represent a serious problem for alignment, we did not apply any filtering method. When sequences and alignments within these non-conserved blocks were checked with the results of the phylogenetic analysis, we noted that the similar sequences in the non-conservative blocks usually define the cluster of species. The exclusion of such information from the phylogenetic analysis would result in average in worse trees and an increase of false support as demonstrated by Tan *et al.* (2015).

The protein coding cox1–3′ and cox1–5′ mtDNA fragments formed matrices of 853 and 659 aligned positions, respectively. The Kimura 2-parameter genetic distances among Agrilus sequences reached up to 22.0, 28.9, and 32.6%, in rrnL, cox1–3′, and cox1–5′ mtDNA, respectively.

Three maximum likelihood (ML) analyses each utilizing a different dataset

Phylogeny was inferred using ML optimality criterion, as implemented in RAxML 8.1 (Stamatakis, 2014). When applicable, each dataset (see below) was partitioned by genes and, alternatively, by gene and codon positions for protein-coding fragments. The nucleotide substitution model applied for all gene fragments was set to GTR + I + G, as suggested by the AIC criterion in jModelTest 2.1.2 (Darriba *et al.*, 2012), with all partitions unlinked. Confidence intervals were determined with 1000 non-parametric bootstrap replicates (Felsenstein, 1985) utilizing the rapid bootstrap option under the GTRGAMMA substitution model. Recovered internal *Agrilus* clades were compared with the species-groups as defined by Jendek & Grebennikov (2011) and by Alexeev (1998, as subgenera).

All three fragment-specific datasets differ notably in the number of terminals. To address this data inequality, three separate analyses, each utilizing separate datasets, were created. *Analysis 1* utilized the densest and least inclusive three-fragment dataset containing 124 terminals (including five non-*Agrilus* outgroups). The average representation was 2.36 fragments per terminal, while no terminal was represented by less than two fragments. In this analysis, all *Agrilus* species (and in some cases geographically distant conspecific populations and/or populations collected from different host plants) were represented by a single terminal. *Analysis 2* utilized the least dense and most inclusive three-fragment dataset containing 475 terminals (including non-*Agrilus* five outgroups), many of them represented by the DNA barcode fragment only (although the majority of species had at least one additional fragment sequenced for at least a single individual) with the average representation 1.64 fragments per terminal. *Analysis 3* utilized the one-fragment (DNA barcode) dataset containing 725 terminals (including 14 *Coraebus* outgroups). For this analysis, our newly generated DNA barcodes were supplemented with those publicly available from the BOLD database (http://www.boldsystems.org, accessed on 13 July

2017, their list is in table S2) and merged in the 475-taxa dataset.

Species delimitation analyses

The morphology-based identification of *Agrilus* specimens resulted in a total of 85 named *Agrilus* species, plus 15 unnamed provisional species. The morphologically defined species originated from Europe and the Mediterranean Region (∼40 spp.), Eastern Asia (∼40 spp.) and North America (∼20 spp.). In some cases, an unidentified specimen clustered on the trees with presumably conspecific individuals with similar DNA sequences; provisional species names for such specimens are given in brackets in fig. S1.

To investigate an algorithm-based species delimitation in *Agrilus*, putative species were estimated alternatively using 2 and 3% DNA barcoding thresholds, as implemented in Species Identifier 1.7.7 (Meier *et al.*, 2006). The morphology-based species definitions were separately tested by using uncorrected pairwise distances among *Agrilus* sequences for all three sequenced fragments. Further, we identified putative species entities by implementing the Bayesian Poisson Tree Processes (bPTP) model for species delimitation (the bPTP server at species.h-its.org). The algorithm delimits putative species from the tree with branch lengths representing the number of substitutions (Zhang *et al.*, 2013) and provides posterior probabilities of all descendants under a node representing a single species. We ran separate analyses for the trees inferred using ML criterion and utilizing three datasets, each constituted by a single DNA fragment (*cox1–3′*, *cox1–5′*, and *rrnL*). The highest numbers of individuals and morphologically pre-defined species were represented in the *cox1–5′* (DNA barcode) dataset with 444 sequences and 94 morphologically delimited species (named or unnamed); the *rrnL* dataset contained 166 sequences and 69 species, while the *cox1–3′* dataset contained 170 sequences and 75 species.

Results

Considering that analysis 1 utilized a pruned dataset with a lesser fraction of missing data (as compared with analysis 2), its topology is reported herein and discussed in more detail, while analysis 2 (fig. S1), although generally congruent, is mentioned only for comparative purposes.

ML analyses of the three-fragment datasets (analyses 1 and 2) recovered monophyletic *Agrilus* with bootstrap support (BS) of 65% (fig. 4) and 60% (fig. S1), respectively. Subsequent basal-most splits within *Agrilus* are conflictingly resolved in both analyses and have negligibly low statistical support with SB often <10%. Both analyses consistently recovered six morphologically pre-defined species-groups (fig. 4), although in most cases these clades did not have a well-supported sister clade (with the exception of the *convexicollis*- and *cyanescens* species-groups, which together formed a strongly supported clade with BS 93%, fig. 4).

The dichotomies corresponding to individual species and small groups of species are often statistically well-supported with BS >70% (fig. 4). Many morphologically pre-defined species were rendered paraphyletic by representatives of other nominal species. Thus, for example, *A. asahinai* from the Russian Far East is nested inside of *A. cyanescens* (fig. 4); the latter is a morphologically coherent species with a native Palaearctic distribution (also introduced to North America). Perhaps the most notable incongruence among the
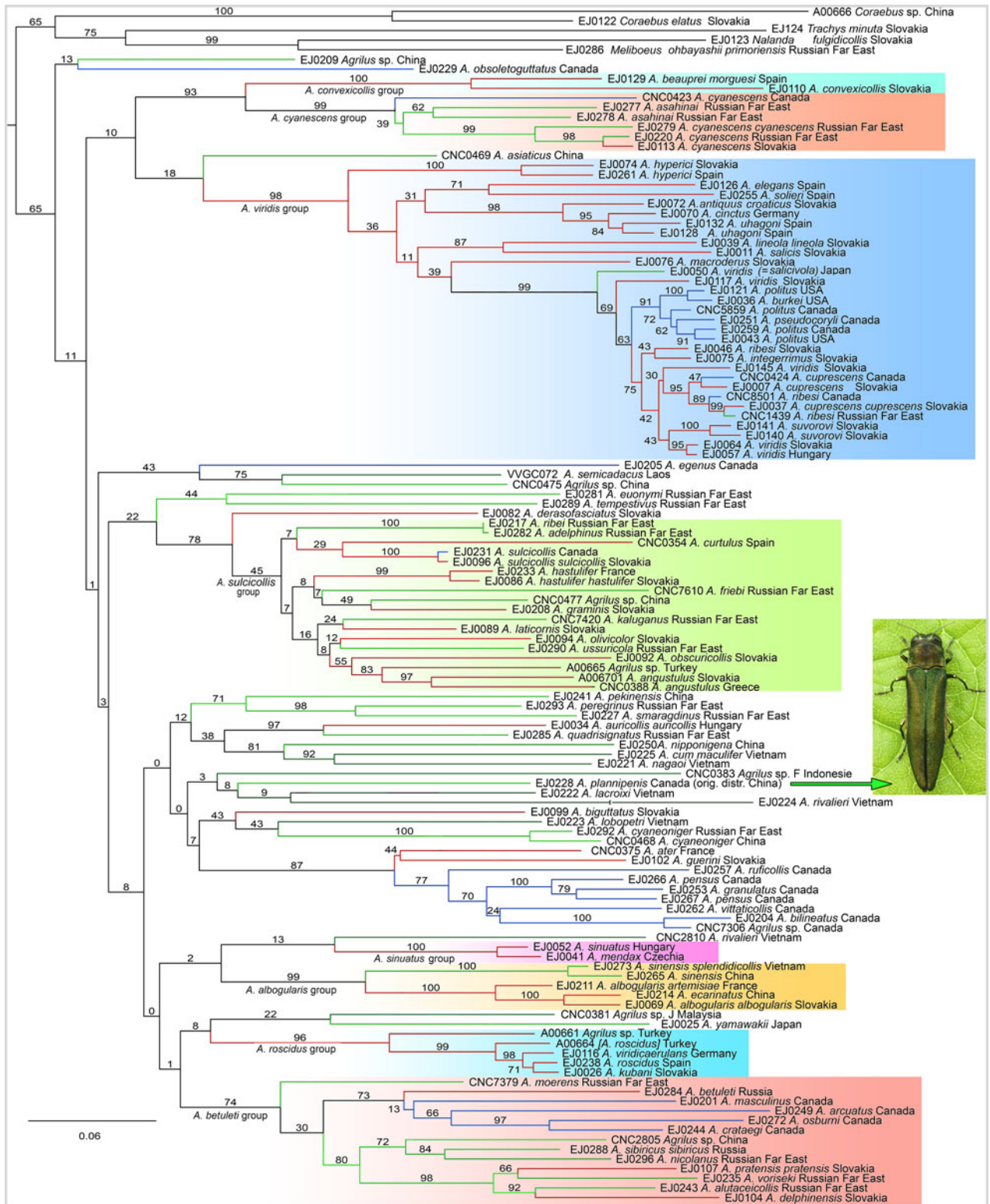
Fig. 4. The phylogram of the genus *Agrilus* recovered by Analysis 1. Numbers at internodes are bootstrap values; species groups are highlighted in different background colors; colors of ingroup branches correspond to geographical units (red: Europe, green: Asia, blue: North America).

Table 1. Comparison of the algorithmic and morphological identifications of *Agrilus*. Details in Supplementary table S3.

Dataset (marker)	Uncorrected pairwise genetic distance						Bayesian PTP		
	rrnL	*rrnL*	*cox1* -Bar	*cox1* -Bar	*cox1–3′*	*cox1–3′*	*rrnL*	*cox1* -Bar	*cox1–3′*
Threshold	3%	2%	3%	2%	3%	2%	–	–	–
Number of putative DNA-based species	65	76	104	115	78	84	79	171	85
Number of morpho-logy-based species	69	69	95	95	75	75	69	95	75

morphologically pre-defined species is depicted by the strongly supported (BS 99%) clade consisting of *A. viridis* (Linnaeus) (specimen EJ0050) and its sister-clade containing, in addition to the same nominal species, specimens of seven other species.

Depending on the method, the number of putative species identified using the *cox1–5′* dataset varied between 104 and 171 (table 1 and fig. S2). The lowest number of putative species was recovered when the 3% uncorrected genetic distance threshold was applied, while the highest number was recovered when the putative species was inferred using bPTP (tables 1 and S3; same species always designated with the same letter of the combination of letters, numbers designate merged subsets of individuals). The number of putative species inferred from the *rrnL* and *cox1–3′* datasets varied between 65 and 79, and between 78 and 85, respectively (tables 1 and S3).

Discussion

Limitations of the analyses

This work is the first and perhaps too bold of an attempt to uncover the internal phylogenetic structure of the largest animal genus by using only three mitochondrial fragments sequenced somewhat inconsistently for ~3% of the known species diversity. The reported results herein, therefore, do not match in their clarity and statistical support those obtained in more thoroughly executed analyses of other mega-diverse animal genera (i.e. Maddison, 2012 on *Bembidion* Laterille; Breeschoten *et al.*, 2016 on *Onthophagus* Latreille). Besides, however, releasing sequence data and thus making them publicly available, our results are informative enough to draw a few preliminary conclusions.

First genus-wide DNA reference library of Holarctic Agrilus

The currently reported analysis is based on inexpensive and rapidly obtainable DNA sequences of three short mitochondrial fragments. The recent efforts to document and quantify diversity using molecular data generally follow two main directions. In the minimalistic approach, the short *cox1–5′* mtDNA fragment (=DNA barcode) is mass-sequenced and deposited in the Barcode of Life Data System, which at present contains data on over 250,000 nominal animal species (barcodeoflife.org; Hebert *et al.*, 2003; Pentinsaari *et al.*, 2014a). Alternatively, phylogenetic trees are normally produced using a higher number of markers (such as the entire mitochondrial genome) and might later be combined with densely sampled barcode databases (Bocak *et al.*, 2014; Zhou *et al.*, 2016; Linard *et al.*, 2018). Here, we combine with satisfactory results the DNA barcode data (444 fragments of *cox1–5′* mtDNA) and two additional mtDNA fragments (166 fragments of *rrnL* and 170 fragments of *cox1–3′* mtDNA) in the single dataset capable of resolving ~100 species, i.e. ~3% of *Agrilus* diversity. Although comparable to the 2.3% density

of DNA samples available for the whole of Coleoptera (Bocak *et al.*, 2014), this sampling density is still scanty, particularly considering the geographical and biological spread of the genus *Agrilus*. With all these limitations, our results are first of their kind for the Holarctic representatives of the genus *Agrilus* and as such are the best available for the tree-based identification of unknown congeneric samples.

Agrilus diversity: species-groups and their relationships

Even though this study gives the first insight in the molecular phylogeny of a high number of *Agrilus* species, the statistical support for the basal-most dichotomies is virtually absent (figs 4 and 5). This is likely due to the DNA saturation effect, especially in the third positions of the protein-coding fragments. The robustly supported clades are few in the number of species usually containing 10 or fewer species. Even with these limitations, our topology (fig. 4) includes all major *Agrilus* species-groups of the Northern temperate zone and we suppose that inclusion in the analysis of an additional sample should, in most cases, lead to its reliable placement among the most closely related organisms (Bocak *et al.*, 2014; Zhou *et al.*, 2016; Linard *et al.*, 2018).

The main *Agrilus* intrageneric lineages detected herein are marked on the tree (fig. 4) and correspond to the taxonomically recognized species-groups (or subgenera), shown in table S4 (Alexeev, 1988; Jendek & Grebennikov, 2011). The sister-group
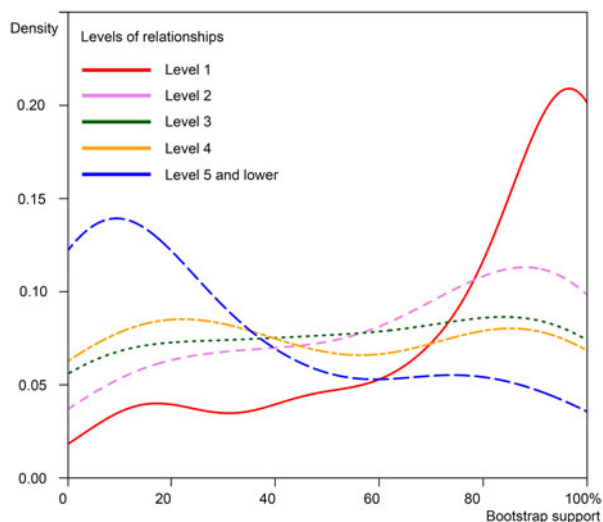

Fig. 5. Densities of bootstrap support for various levels of relationships in the phylogram depicted in fig. 4. Level 1 represents the sister-pairs; level 2 represents the sister-pair and an additional single species or another sister pair; further levels represent a clade of the previous level plus an additional species or a clade of the same or lower level.

relationship between the *convexicollis*- and *cyanescens* species-groups agrees well with the expectations based on the study of adult morphology. The *viridis* species-group (fig. 4) corresponds to three species-groups defined by Jendek & Grebennikov (2011) and includes widely distributed Palaearctic or Nearctic xylophagous species attacking pioneer trees like *Salix, Populus, Betula, Alnus, Rubus,* and *Corylus.* Nearctic members of this group (*A. politus* and *A. burkei*) are biologically most similar to the Euro-Siberian *A. viridis* and have likely dispersed from Asia to North America, as many other herbivorous beetles (Sota *et al.*, 2008). The *sulcicollis* species-group is morphologically well defined and contains many species in the Palaearctic and Oriental regions developing on *Quercus.* This group is further distinct by its members having pronounced sexual dimorphism. The clade of the *albogularis* species-group is also well defined ecologically and morphologically comprising rhizophagous species of desert and semi-desert areas developing on Compositae, Amaranthaceae and Nitrariaceae. It is, therefore, fair to say that the herein recovered topology (fig. 4) presents the first but still very imperfect glimpse on the internal phylogenetic structure of the North Hemisphere *Agrilus.*

Limits of Agrilus species identification

The algorithmic DNA-based delimitation of species was proposed either as an identification tool when the taxonomic identity of unknown sample is inferred from the genetic distance to the identified reference samples (Hebert *et al.*, 2003; www.bold.org), or as a tool for identification of putative species for further studies when biological diversity is evaluated, but where the taxonomical identification is impossible or cost ineffective (Pons *et al.*, 2006; Vogler & Monaghan, 2007; Monaghan *et al.*, 2009). To develop such tools to the genus *Agrilus*, we compared two methods: (1) comparison of uncorrected pairwise distances (Meier *et al.*, 2006) and (2) the species delineating by application of the Bayesian Poisson Tree Process (bPTP; Zhang *et al.*, 2013).

The 3% genetic divergence was often considered as a limit for the delineation of a putative species using the barcoding marker (Hebert *et al.*, 2003; Blaxter, 2004; Smith *et al.*, 2006); however, limits as low as 1% difference were also proposed (Ratnasingham & Hebert, 2007). We tested 2 and 3% thresholds and found that the levels of agreement between morphology and distance-based delimitations varied with individual species clades. The methods using genetic divergence generally overestimated the number of species. The algorithmic analysis of the DNA barcode dataset from 95 morphologically identified species indicated the presence of 104 and 115 species, using 3 and 2% threshold, respectively (table 1). Application of these thresholds for other mtDNA fragments generally estimates a lower species number, which is likely due to the lower genetic variability of these fragments (tables 1, S3).

Similarly, the bPTP analysis generally overestimated the number of species (tables 1, S3). The bPTP tends to over-split species-level clades into a higher number of putative species and, therefore, the analysis of the barcode dataset indicated the presence of 171 species, compared with 95 identified using traditional morphology. With the application of both methods, conspecific populations having distant geographic origin were identified as separate putative species, similar to the results of other authors (e.g., Bergsten *et al.*, 2012; Li *et al.*, 2015). The dense sampling of populations from the whole range is the only way to identify cases where geographic distance is responsible for increased genetic divergence within a population group that is still connected by gene flow. Concerning our results, we cannot recommend a single threshold for species delimitation in *Agrilus.*

We found numerous method-dependant disagreements in the species identifications. Several species were represented by multiple unrelated clades. Although we cannot exclude misidentifications and introgressions (Pentinsaari *et al.*, 2014b), the recovered relationships might at least in some cases indicate hidden diversity. For example, *A. viridis* reared from *Tillia* (fig. S1) represents a distant lineage in regard to the rest of this species. Such a distant position was inferred already by previous authors (Pentinsaari *et al.*, 2014b) and ascribed to introgression. As various distant populations of *A. viridis* from *Tillia* show genetic diversification, cryptic diversity might be the case here. The current concept of *A. viridis* makes this taxon the most polyphagous assemblage with 49 species-level host plants (Jendek & Poláková, 2014). Additionally, *A. viridis* has an extremely large geographic range and various populations of adults active in different periods from April to September and in altitudes ranging from sea level to about 1600 m (Jendek & Grebennikov, 2011). All this suggests that this nominal species might perhaps be better split into many. Prior to any taxonomic actions, however, much denser sampling should be analysed. Another example is *A. ribesi*, which is represented by three distantly related populations (fig. 4): one from Slovakia, one from the Russian Far East (which might represent an introduced population from Europe), and one from Canada, which was reported as an invasive species (Jendek *et al.*, 2015). The third example is formed by two terminals of *A. angustulus* from Greece (fig. S1), representing generically dissimilar lineages and separate species. Both genetic introgression and underestimated species diversity might account for the observed results.

Another common case of a conflict between DNA-based species limits and morphologically defined species are clades containing a single species with genetic divergence markedly higher than one might expect in a group of interbreeding populations with an effective gene flow. These include *A. cyanescens* with several genetically distinct populations from Russia, Canada and Central Europe (figs 4m S1, table 1), which are genetically very distant (17.3 and 8.1% in *cox1–5′* and *cox1–3′*, respectively). The geographic distance precluding effective gene flow is one possible explanation; however, a much denser sampling of Siberian populations is needed to test this hypothesis. Two populations of *A. hyperici* from the western Mediterranean and from Central Europe are also genetically quite distant to each other (7.9 and 6.3% in *cox1–5′* and *cox1–3′*, respectively) and they were recovered as separate species in algorithmic delimitations (table S). Also *A. uhagoni* Abeille de Perrin consists of two notably dissimilar lineages, both developing on *Genista* and both from Spain (fig. 4). Either the genetically diversified populations of *A. uhagoni* with an independent evolutionary history may sympatrically co-occur in Spain, or these samples are misidentified due to the variability of the diagnostic morphological characters. High genetic diversity was also identified in Canadian and US populations of *A. politus* (Say) (fig. 4), but without any sign of the geographic structure. Their maximum genetic distance reached 5.5, 5.2 and 4.4%, in *cox1–5′*, *cox1–3′*, and *rrnL*, respectively. A few other similar examples might be detected on our topologies (such as *A. albogularis* Gory, *A. roscidus* Kiesenwetter, *A. cyaneoniger* Thomson, figs 4, S1, S2).

Many of the above-mentioned species were recovered as paraphyletic in our analyses by having other morphologically delimited species nested within them. A delayed lineage sorting is a possible explanation for the observed conflicts between morphology- and DNA-based species delimitation. The species rendering other species paraphyletic include: (1) *A. antiquus croaticus* Abeille de Perrin and *A. cinctus* (Olivier) nested within *A. uhagoni*; (2) *A. pseudocoryli* Fisher and *A. burkei* Fisher nested within *A. politus*; (3) *A. suvorovi* Obenberger nested within *A. viridis*; (4) *A. perisi* Cobos nested within *A. angustulus*; (5) *A. viscivorus* Bílý and *A. kubani* Bílý nested within *A. roscidus*; (6) *A. pensus* Horn nested within *A. anxius* Gory; (7) *A. ecarinatus* Marseul nested within *A. albogularis*. Many of these cases might suggest imperfect taxonomy. The taxonomic status of *A. burkei* is unclear and sometimes it is considered intraspecific with *A. politus*. The only character to distinguish *A. ecarinatus* is its small body size. We can hypothesize that *A. kubani* and *A. viscivorus* are in fact independent species having different host plants, *Loranthus europaeus* and *Viscum album*, respectively. In other cases we are unable to account for the observed conflict and, therefore, they require further investigations.

Concluding remarks

The comprehensive phylogeny of *Agrilus* remains unattainable due to the extreme diversity of this clade, practical rareness of species, inaccessibility of many regions and the shortage of trained human capacity to document such enormous biological diversity. Nevertheless, the herein released DNA database of several hundred samples and about 100 species provides the first dataset likely capable to identify an unknown *Agrilus* specimen from the northern temperate zone of the Globe. This has significant practical utility, since considering the history of *Agrilus* invasions (Hoebeke *et al.*, 2017), this region contains the highest number of economically important and potentially harmful species. Our results convincingly demonstrate that, unlike the minimalistic DNA barcode approach, the multi-marker phylogenetic analysis has a notably higher potential to place an unknown sample among the phylogenetically closest and taxonomically identified relatives. Although morphological misidentifications and sequencing errors cannot be excluded, the relatively high number of disagreement between morphology- and DNA-based species limits calls for closer and more detailed investigation of such cases. Overall, and at the very least, our work is the first attempt to tackle the phylogenetic structure of what currently is the largest conventionally accepted genus of organisms on Earth, the *Agrilus* jewel beetles.

Supplementary material

The supplementary material for this article can be found at https://doi.org/10.1017/S0007485318000330

References

Ahrens, D., Monaghan, M.T. & Vogler, A.P. (2007) DNA-based taxonomy for associating adults and larvae in multi-species assemblages of chafers (Coleoptera: Scarabaeidae). *Molecular Phylogenetics & Evolution* **44**, 436–449. doi: 10.1016/j.ympev.2007.02.024.

Akers, R.C., Herms, D.A. & Nielsen, D.G. (1986) Emergence and adult biology of *Agrilus difficilis* (Coleoptera: Buprestidae), a pest of honeylocust, *Gleditsia triacanthos*. *Great Lakes Entomologist* **19**, 27–30.

Alexeev, A.V. (1998) Towards subgeneric classification of *Agrilus* curtis jewel beetles (Coleoptera: Buprestidae) of the Palaearctic fauna. *Entomologicheskoe Obozrenie* **77**, 367–383. [in Russian].

Ashfaq, M. & Hebert, P.D.N. (2016) DNA barcodes for bio-surveilance: regulated and economically important arthropod plant pests. *Genome* **59**, 933–945.

Aukema, J.E., McCullough, D.G., Von Holle, B., Liebhold, A.M., Britton, K. & Frankel, S.J. (2010) Historical accumulation of nonindigenous forest pests in the continental US. *BioScience* **60**, 886–897.

Aukema, J.E., Leung, B., Kovacs, K., Chivers, C., Britton, K.O., Englin, J., Frankel, S.J., Haight, R.G., Holmes, T.P., Liebhold, A.M., McCullough, D.G. & Von Holle, B. (2011) Economic impacts of non-native forest insects in the continental United States. *PLoS ONE* **6**(9), e24587. doi: 10.1371/journal.pone.0024587.

Baselga, A., Gomez-Rodriguez, C., Novoa, F. & Vogler, A.P. (2013) Rare failures of DNA bar codes to separate morphologically distinct species in a biodiversity survey of Iberian leaf beetles. *PLoS ONE* **8**, e75854.

Bellamy, C.L. (2008) *A World Catalogue and Bibliography of the Jewel Beetles (Coleoptera: Buprestoidea)*, Volume **4**, Agrilinae: Agrilina through Trachyini. Sofia, Pensoft, 1932–2684 pp.

Bergsten, J., Bilton, D.T., Fujisawa, T., Elliott, M., Monaghan, M. T., Balke, M., Heindrich, L., Geijer, J., Herrmann, J., Foster, G.N., Ribera, I., Nulsson, A.N., Barraclough, T.G. & Vogler, A.P. (2012) The effect of geographical scale of sampling on DNA barcoding. *Systematic Biology* **61**, 851–869.

Blaxter, M.L. (2004) The promise of a DNA taxonomy. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **359**, 669–679.

Bocak, L., Bocakova, M., Hunt, T. & Vogler, A.P. (2008) Multiple ancient origins of neoteny in lycidae (Coleoptera): consequences for ecology and macroevolution. *Proceedings of the Royal Society B, Biological Sciences* **275**, 2015–2023. doi: 10.1098/rspb.2008.0476.

Bocak, L., Barton, C., Crampton-Platt, A., Chesters, D., Ahrens, D. & Vogler, A.P. (2014) Building the Coleoptera tree-of-life for >8000 species: composition of public DNA data and fit with Linnaean classification. *Systematic Entomology* **39**, 97–110.

Bocek, M. & Bocak, L. (2016) Where are species limits in polymorphic mimetic beetles from New Guinean mountains: a case of *Eniclases* net-winged beetles (Lycidae: Metriorrhynchini). *ZooKeys* **593**, 15–35. doi: 10.3897/zookeys.593.7728.

Breeschoten, T., Doorenweerd, C., Tarasov, S. & Vogler, A.P. (2016) Phylogenetics and biogeography of the dung beetle

genus *Onthophagus* inferred from mitochondrial genomes. *Molecular Phylogenetics and Evolution* **105**, 86–95. doi: 10.1016/j.ympev.2016.08.016.

Carstens, B.C., Pelletier, T.A., Reid, N.M. & Satler, J.D. (2013) How to fail at species delimitation. *Molecular Ecology* **22**, 4369–4383. doi: 10.1111/mec.12413.

Darriba, D., Taboada, G.L., Doallo, R. & Posada, D. (2012) Jmodeltest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**, 772.

Drès, M. & Mallet, J. (2002) Host races in plant-feeding insects and their importance in sympatric speciation. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences* **357**(1420), 471–492. doi: 10.1098/rstb.2002.1059.

Dunn, J.P., Kimmerer, T.W. & Nordin, G.L. (1986) The role of host tree condition in attack of white oaks by the twolined chestnut borer, *Agrilus bilineatus* (Coleoptera: Buprestidae). *Oecologia* **70**, 596–600.

Egan, S.P., Nosil, P. & Funk, D.J. (2008) Selection and genomic differentiation during ecological speciation: isolating the contributions of host association via a comparative genome scan of *Neochlamisus bebbianae* leaf beetles. *Evolution* **62**, 1162–1181. doi: 10.1111/j.1558-5646.2008.00352.x.

Everett, R.A. (2000) Patterns and pathways of biological invasions. *Trends in Ecology & Evolution* **15**, 177–178.

Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791. doi: 10.2307/2408678.

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.

Gibbs, J.N. & Greig, B.J.W. (1997) Biotic and abiotic factors affecting the dying back of pedunculate oak *Quercus robur* L. *Forestry* **70**, 399–406.

Gordon, S.C., Woodford, J.A.T. & Birch, A.N.E. (1997) Arthropod pests of *Rubus* in Europe: pest status, current and future control strategies. *Journal of Horticultural Science* **72**, 831–862.

Grebennikov, V.V., Jendek, E. & Smirnov, M.Ed. (2017) Diagnostic and phylogenetic utility of the first DNA barcode library for longhorn beetles (Coleoptera: Cerambycidae) from the Russian Far East. *Zootaxa* **4276**, 441–445. doi: 10.11646/zootaxa.4276.3.9.

Haack, R.A., Benjamin, D.M. & Haack, K.D. (1983) Buprestidae, Cerambycidae, and Scolytidae associated with successive stages of *Agrilus bilineatus* (Coleoptera: Buprestidae) infestation of oaks in Wisconsin. *Great Lakes Entomologist* **16**, 47–55.

Haack, R.A., Jendek, E., Liu, H., Marchant, K.R., Petrice, T.R., Poland, T.M. & Ye, H. (2002) The emerald ash borer: a new exotic pest in North America. *Newsletter of the Michigan Entomological Society* **47**(3–4), 1–5.

Hebert, P.D.N., Ratnasingham, S. & deWaard, J.R. (2003) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London B: Biological Sciences* **270**, 96–99.

Herms, D.A. & McCullough, D.G. (2014) Emerald Ash borer invasion of North America: history, biology, ecology, impacts, and management. *Annual Review of Entomology* **59**, 13–30.

Hoebeke, E.R., Jendek, E., Zablotny, J.E., Rieder, R., Yoo, R., Grebennikov, V.V. & Ren, L. (2017) First North American records of the east Asian metallic wood-boring beetle *Agrilus smaragdifrons* Ganglbauer (Coleoptera: Buprestidae: Agrilinae), a specialist on tree of heaven (*Ailanthus altissima*,

Simaroubaceae). *Proceedings of the Entomological Society of Washington* **119**, 408–422. doi: 10.4289/0013-8797.119.3.408.

Ivanova, N.V., de Waard, J.R. & Hebert, P.D.N. (2006) An inexpensive, automation-friendly protocol for recovering high quality. DNA. *Molecular Ecology Notes* **6**, 998–1002. doi: 10.1111/j.1471-8286.2006.01428.x.

Jendek, E. (2016) Taxonomic, nomenclatural, distributional and biological study of the genus *Agrilus* (Coleoptera: Buprestidae). *Journal of Insect Biodiversity* **4**(2), 1–57.

Jendek, E. & Grebennikov, V. (2009) *Agrilus sulcicollis* (Coleoptera: Buprestidae) a new alien species in North America. *Canadian Entomologist* **141**, 236–245.

Jendek, E. & Grebennikov, V. (2011) *Agrilus (Coleoptera, Buprestidae) of East Asia*. Prague, Jan Farkač, 362 pp.

Jendek, E. & Poláková, J. (2014) *Host Plants of World Agrilus (Coleoptera, Buprestidae). A Critical Review*. Springer International Publishing AG, Cham, Springer Cham Heidelberg New York Dordrecht London, 706 pp.

Jendek, E., Grebennikov, V. & Bocak, L. (2015) Undetected for a century: Palaearctic *Agrilus ribesi* Schaefer (Coleoptera: Buprestidae) on currant in North America, with adult morphology, larval biology and DNA barcode. *Zootaxa* **4034**, 112–126.

Jones, E.A., Reed, D.D., Mroz, G.D., Liechty, H.O. & Cattelino, P.J. (1993) Climate stress as a precursor to forest decline: paper birch in northern Michigan, 1985–1990. *Canadian Journal of Forest Research* **23**, 229–233. doi: 10.1139/x93-030.

Kang, A.R., Kim, M.J., Park, I.A., Kim, K.Y. & Kim, I. (2016) Extent and divergence of heteroplasmy of the DNA barcoding region in *Anapodisma miramae* (Orthoptera: Acrididae). *Mitochondrial DNA Part A* **27**, 3405–3414. doi: 10.3109/19401736.2015.1022730.

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

Kovacs, K.F., Mercader, R.J., Haight, R.G., Siegert, N.W., McCullough, D.G. & Liebhold, A.M. (2011) The influence of satellite populations of emerald ash borer on projected economic costs in U.S. Communities, 2010–2020. *Journal of Environmental Management* **92**, 2170–2181.

Kusy, D., Sklenarova, K. & Bocak, L. (2018) The effectiveness of DNA-based delimitation in *Synchonnus* net-winged beetles (Coleoptera: Lycidae) assessed, and description of 11 new species. *Austral Entomology* **57**, 25–39.

Li, Y., Gunter, N., Pang, H. & Bocak, L. (2015) DNA-based species delimitation separates highly divergent populations within morphologically coherent clades of poorly dispersing beetles. *Zoological Journal of the Linnean Society* **175**, 59–72.

Linard, B., Crampton-Platt, A., Moriniere, J., Timmermans, M.J.T.N., Andujar, C., Arribas, P., Miller, K.E., Lipecki, J., Favreau, E., Hunter, A., Gomez-Rodriguez, C., Barton, C., Nie, R., Gillett, C.P.D.T., Breeschoten, T., Bocak, L. & Vogler, A.P. (2018) The contribution of mitochondrial metagenomics to largescale data mining and phylogenetic analysis of Coleoptera. bioRxiv 280792. doi: 10.1101/280792.

Liu, H.P., Bauer, L.S., Gao, R.T., Zhao, T.H., Petrice, T.R. & Haack, R.A. (2003) Exploratory survey for the emerald ash borer, *Agrilus planipennis* (Coleoptera: Buprestidae), and its natural enemies in China. *Great Lakes Entomologist* **36**, 191–204.

Maddison, D.R. (2012) Phylogeny of *Bembidion* and related ground beetles (Coleoptera: Carabidae: Trechinae: Bembidiini: Bembidiina). *Molecular Phylogenetics and Evolution* **63**, 533–576.

Meier, R., Shiyang, K., Vaidya, G. & Ng, P.K.L. (2006) DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology* **55**, 715–728.

Monaghan, M.T., Wild, R., Elliot, M., Fujisawa, T., Balke, M., Inward, D.J., Lees, D.C., Ranaivosolo, R., Eggleton, P., Barraclough, T.G. & Vogler, A.P. (2009) Accelerated species inventory on Madagascar using coalescent-based models of species delimitation. *Systematic Biology* **58**, 298–311.

Nosil, P. & Mooers, A.O. (2005) Testing hypotheses about ecological specialization using phylogenetic trees. *Evolution* **59**, 2256–2263.

Pentinsaari, M., Hebert, P.D.N. & Mutanen, M. (2014*a*) Barcoding beetles: a regional survey of 1872 species reveals high identification success and unusually deep interspecific divergences. *PLoS ONE* **9**(9), e108651. doi: 10.1371/journal.pone.0108651.

Pentinsaari, M., Mutanen, M. & Kaila, L. (2014*b*) Cryptic diversity and signs of mitochondrial introgression in the *Agrilus viridis* species complex (Coleoptera: Buprestidae). *European Journal of Entomology* **111**, 475–486. doi: 10.14411/eje.2014.072.

Pons, J., Barraclough, T.G., Gomez-Zurita, J., Cardoso, A., Duran, D.P., Hazell, S., Kamoun, S., Sumlin, W.D. & Vogler, A.P. (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology* **55**, 595–609.

Ratnasingham, S. & Hebert, P.D.N. (2007) BOLD: the barcode of life data system. *Molecular Ecology Notes* **7**, 355–364.

Riedel, A., Sagata, K., Surbakti, S., Rene, T. & Balke, M. (2013a) One hundred and one new species of *Trigonopterus* weevils from New Guinea. *ZooKeys*, **280**, 1–150.

Riedel, A., Sagata, K., Suhardjono, Y.R., Tänzler, R. & Balke, M. (2013b) Integrative taxonomy on the fast track – towards more sustainability in biodiversity research. *Frontiers in Zoology* **10**, 15.

Satler, J.D., Carstens, B.C. & Hedin, M. (2013) Multilocus species delimitation in a complex of morphologically conserved trapdoor spiders (Mygalomorphae, Antrodiaetidae, Aliatypus). *Systematic Biology* **62**, 805–823. doi: 10.1093/sysbio/syt041.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H. & Flook, P. (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America* **87**, 651–701.

Smith, M.A., Woodley, N.E., Janzen, D.H., Hallwachs, W. & Hebert, P.D.N. (2006) DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). *Proceedings of the National Academy of Sciences of the United States of America* **103**, 3657–3662.

Sota, T., Bocak, L. & Hayashi, M. (2008) Molecular phylogeny and historical biogeography of the Holarctic wetland leaf beetle of the genus *Plateumaris*. *Molecular Phylogenetics and Evolution* **46**, 183–192.

Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics (Oxford, England)* **30**, 1312–1313.

Tan, G., Muffato, M., Ledergerber, C., Herrero, J., Goldman, N., Gil, M. & Dessimoz, C. (2015) Current methods for automated filtering of multiple sequence alignments frequently worsen single-gene phylogenetic inference. *Systematic Biology* **64**, 778–791. doi: 10.1093/sysbio/syv033.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. & Higgins, D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**, 4876–4882.

Vansteenkiste, D., Tirry, L., Van Acker, J. & Stevens, M. (2004) Predispositions and symptoms of *Agrilus* borer attack in declining oak trees. *Annals of Forest Science* **61**, 815–823. doi: 10.1051/forest:2004076.

Vences, M., Guayasamin, J.M., Miralles, A. & De la Riva, I. (2013) To name or not to name: criteria to promote economy of change in Linnaean classification schemes. *Zootaxa* **3636**, 201–244.

Vogler, A.P. & Monaghan, M.T. (2007) Recent advances in DNA taxonomy. *Journal of Zoological Systematics and Evolutionary Research* **45**, 1–10.

Wargo, P.M. (1977) *Armillariella mellea* and *Agrilus bilineatus* and mortality of defoliated oak trees. *Forest Science* **23**, 485–492.

Zahiri, R., Lafontaine, J.D., Schmidt, B.C., de Waard, J.R., Zakharov, E.V. & Hebert, P.D.N. (2014) A transcontinental challenge – A test of DNA barcode performance for 1,541 species of Canadian Noctuoidea (Lepidoptera). *PLoS ONE* **9**, e92797.

Zhang, J.J., Kapli, P., Pavlidis, P. & Stamatakis, A. (2013) A general species delimitation method with applications to phylogenetic placements. *Bioinformatics (Oxford, England)* **29**, 2869–2876. doi: 10.1093/bioinformatics/btt499.

Zhou, X., Frandsen, P.B., Holzenthal, R.W., Beet, C.R., Bennett, K.R., Blahnik, R.J., Bonada, N., Cartwright, D., Chuluunbat, S., Cocks, G.V., Collins, G.E., de Waard, J.R., Dean, J., Flint, O.S., Hausmann, A., Hendrich, L., Hess, M., Hogg, I.D., Kondratieff, B.C., Malicky, H., Milton, M.A., Morinière, J., Morse, J.C., Mwangi, F.N., Pauls, S.U., Gonzalez, M.R., Rinne, A., Robinson, J.L., Salokannel, J., Shackleton, M., Smith, B., Stamatakis, A., StClair, R., Thomas, J.A., Zamora-Muñoz, C., Ziesmann, T. & Kjer, K.M. (2016) The Trichoptera barcode initiative: a strategy for generating a species-level tree of life. *Proceedings of the Royal Society of London B: Biological Sciences* **371**, 20160025. doi: 10.1098/rstb.2016.0025.