Comparison of five methods for delimitating species in Ophion Fabricius, a diverse genus of parasitoid wasps (Hymenoptera, Ichneumonidae)

Marla D. Schwarzfeld *, Felix A.H. Sperling

Department of Biological Sciences, CW 405 Biological Sciences Building, University of Alberta, Edmonton, Alberta T6G 2E9, Canada

ABSTRACT

DNA taxonomy has been proposed as a method to quickly assess diversity and species limits in highly diverse, understudied taxa. Here we use five methods for species delimitation and two genetic markers (COI and ITS2) to assess species diversity within the parasitoid genus, Ophion. We searched for compensatory base changes (CBCs) in ITS2, and determined that they are too rare to be of practical use in delimiting species in this genus. The other four methods used both COI and ITS2, and included distance-based (threshold analysis and ABGD) and tree-based (GMYC and PTP) models. We compared the results of these analyses to each other under various parameters and tested their performance with respect to 11 Nearctic species/morphospecies and 15 described Palearctic species. We also computed barcode accumulation curves of COI sequences to assess the completeness of sampling. The species count was highly variable depending on the method and parameters used, ranging from 47 to 168 species, with more conservative estimates of 89–121 species. Despite this range, many of the Nearctic test species were fairly robust with respect to method. We concluded that while there was often good congruence between methods, GMYC and PTP were less reliant on arbitrary parameters than the other two methods and more easily applied to genetic markers other than COI. However, PTP was less successful at delimiting test species than was GMYC. All methods, as well as the barcode accumulation curves, indicate that several Palearctic species remain undescribed and that we have scarcely begun to appreciate the Nearctic diversity within this genus.

1. Introduction

Species are a basic unit in studies of ecology, biogeography, and evolution, and are essential to applications in conservation or biological control (Claridge et al., 1997; Sites and Marshall, 2004). But most of Earth’s species remain undescribed, while taxonomists are too few and traditional taxonomy is too slow to deal fully with this unknown diversity (Brooks and Hoheg, 2000; Godfray, 2002; Wheeler, 2004). With the advent of rapid DNA sequencing, molecular taxonomy has been proposed for quickly assessing species diversity in diverse, little-known taxa (Blaxter, 2004; Pons et al., 2006; Tautz et al., 2003). In particular, the Barcode of Life project has popularized use of half of the cytochrome oxidase I (COI) gene of mitochondrial DNA as a standardized “barcode” for species identification and discovery (Hebert et al., 2003a,b; Miller, 2007).

Many studies using DNA taxonomy, particularly of the COI gene, have delimited species with distance-based clustering methods. For example, species have been defined as terminal monophyletic clusters on neighbor-joining trees (Hajibabaei et al., 2006), by a standard interspecific threshold of 1–3% (Hebert et al., 2003a,b, 2004; Smith et al., 2009, 2013; Stalhut et al., 2013; Strutzenberger et al., 2011; Tang et al., 2012), or by an interspecific distance of 10X the intraspecific distance (Hebert et al., 2004). While it is widely acknowledged that a single threshold divergence will not apply to all taxa (Cognato, 2006; Hendrich et al., 2010; Monaghan et al., 2009), pre-defined thresholds have nonetheless been implicitly or explicitly used as criteria for assessing species diversity in a number of studies (Barrett and Hebert, 2005; Hebert et al., 2003a; Smith et al., 2009, 2013; Strutzenberger et al., 2011). As well, the universality of barcoding gaps is controversial, with several studies finding significant overlap between intra- and interspecific divergences (Meyer and Paulay, 2005; Wiemers and Fiedler, 2007). The challenge of objectively determining this gap led to the development of the automated barcode gap discovery (ABGD) algorithm, which infers confidence intervals for intraspecific divergences, and recursively partitions the sequences...
according to the empirically observed divergences (Puillandre et al., 2012a).

Distance-based methods have been criticized for being purely phenetic, rather than incorporating evolutionary theory (DeSalle, 2007; Rubinoff et al., 2006; Will and Rubinoff, 2004; Will et al., 2005). Some explicitly evolutionary approaches have been developed to address this concern, such as coalescence or tree-based methods, and these may be used to delimit large numbers of species (Carstens and Dewey, 2010; Fujita et al., 2012; Leaché and Fujita, 2010; O’Meara, 2010; Yang and Rannala, 2010). One of these methods, the generalized mixed Yule coalescent (GMYC) model, determines the point at which coalescent branching patterns (within species) transition to Yule patterns (between species) (Fontaneto et al., 2007; Pons et al., 2006). This method has shown promise for identifying species boundaries in diverse, little known taxa (Astrin et al., 2012; Cecarelli et al., 2012; Monaghan et al., 2009); however, the results depend on the choice of parameters used to construct the ultrametric tree (Cecarelli et al., 2012). The Poisson Tree Processes model (PTP) is another tree-based method that uses a non-ultrametric phylogenetic tree as input, thus avoiding the challenges of constructing an accurate ultrametric tree (Zhang et al., 2013). The method delimits species under the assumption that the number of substitutions within a species is significantly lower than the number of substitutions between species.

Regardless of the delimitation method used, reliance on a single genetic marker may give misleading results (Dupuis et al., 2012). While COI has many advantages for studies of species delimitation, such as its ease of amplification and relatively rapid rate of mutation (Li et al., 2010; Monti et al., 2005; Williams et al., 2012), it may not accurately delimit species due to factors such as introgression, retained ancestral polymorphisms, or coamplification of nuclear pseudogenes (numts) (Cognato, 2006; Dupuis et al., 2012; Funk and Omland, 2003; Meier et al., 2006; Schmidt and Sperling, 2008). It is therefore important to use additional data, such as from other genes or morphology, in an integrative taxonomic analysis to more accurately delimit species (Dayrat, 2005; Roe and Sperling, 2007; Schlick-Steiner et al., 2010; Schwarzfeld and Sperling, 2014).

The internal transcribed spacer 2 (ITS2) of ribosomal RNA is another rapidly evolving genetic marker that has proved useful for phylogenetic analyses at the species level (Alvarez and Hoy, 2002; Campbell et al., 1993; Gomez-Zurita et al., 2000;Hung et al., 2004; Li et al., 2010; Wagener et al., 2006). Ribosomal RNA is found in multiple tandem repeats within the genome, which undergo concerted evolution so that (in most taxa) the copies are identical (Hillis and Davis, 1988; Ji et al., 2003; but see Keller et al., 2006). Because of these multiple copies, ITS2 is relatively easy to amplify. However, the high rate of mutation and the presence of numerous insertion-deletion events (indels) can make accurate alignment of this gene challenging, particularly among more distantly related species (Coleman and Vacquier, 2002). ITS2 has a highly conserved secondary structure, with characteristic paired domains and unpaired regions (Coleman, 2007). Incorporating this secondary structure can improve alignment of ITS2, thus facilitating phylogenetic reconstruction, and also permits the discovery of compensatory base changes (CBCs)’ (Coleman and Vacquier, 2002; Wiemers et al., 2009; Wolf et al., 2013). These are substitutions where two paired nucleotides (often quite far apart in the linear sequence) both mutate such that the original pairing is maintained (Coleman, 2009; Coleman and Vacquier, 2002; Gettel et al., 1994). Müller et al. (2007) determined that even a single CBC between two individuals indicates that enough evolutionary time has passed for reproductive isolation to arise, resulting in separate species status. Such CBCs do not cause reproductive isolation directly, instead simply indicating a correlation (Wolf et al., 2013). As well, a lack of CBCs is not proof of conspecificity, as closely related species may not have any CBCs (Müller et al., 2007; Wiemers et al., 2009). CBC’s therefore only indicate a minimum number of species among the specimens examined.

**Ophion** is a genus of large-sized (10–20 mm) nocturnal parasitoid wasps in the subfamily Ophioninae of the family Ichneumonidae. Unlike the majority of ophionine genera, **Ophion** are most diverse in temperate regions, with few species in tropical habitats (Gauld, 1980, 1988). **Ophion** are large, abundant, and come readily to lights; however, Neartic species have received almost no taxonomic attention despite being highly apparent and easily collected. Seventeen Neartic species are currently described (Schwarzfeld and Sperling, 2014; Yu et al., 2012), yet it has been estimated that the fauna consists of approximately 50 species (Gauld, 1985). In the Palearctic region, **Ophion** are better known, with 79 described species (Yu et al., 2012), but there have been no species revisions that span the Palearctic. The fauna has been most comprehensively examined in the United Kingdom, with a number of revisions (Gauld, 1973, 1976, 1978) culminating in a comprehensive revision of 16 known species (Brock, 1982).

**Ophion** are notoriously difficult to identify as they are morphologically homogenous among species and yet variable within species (Brock, 1982; Gauld, 1980). They are thus well-suited to the application of molecular taxonomic methods as a first step toward species delimitation. Here we estimate **Ophion** species diversity using tree-based methods (GMYC, PTP), distance-based methods (ABGD, threshold analysis), and by assessing compensatory base changes of ITS2. The study is mainly based on Canadian specimens, with the addition of specimens from the United States and representatives of all species, except one, that have been described from the United Kingdom.

### 2. Methods

#### 2.1. Specimen collection

A total of 674 specimens of **Ophion** were included in this study, including 572 from Canada, 81 from Europe, 19 from the United States, and 2 from Costa Rica (Supplementary Table 1). Most Nearctic specimens were newly collected and sequenced for this study, while 25 were sequenced from alcohol-preserved material on loan from the Canadian National Collection of Insects, Arachnids and Nematodes, and 92 COI sequences from British Columbia were included from the Barcode of Life database (BOLD), courtesy of J. DeWaard (project code: ICHBC). Specimens selected for sequencing represent the range of morphological variation observed in over 4000 specimens from a variety of habitats. Eighty specimens from 15 species of Palearctic **Ophion** were also newly sequenced, all of which are from the United Kingdom except for one specimen from Spain and one from France. Finally, one additional sequence of **O. obscursatus** Fabricius (also from the UK) and two sequences of **O. flavius** Brullé from Costa Rica were included from Genbank and BOLD. With the exception of sequences obtained from GenBank or BOLD, all Neartic taxa used in the current study were identified by M. D. Schwarzfeld. Palearctic taxa were identified by G. R. Broad (Natural History Museum, London), with the exception of **O. forticornis**, which was identified by M. R. Shaw (National Museums Scotland, Edinburgh).

#### 2.2. Sequencing

DNA was extracted from a single hind leg and amplified following methods described in Schwarzfeld and Sperling (2014). We sequenced COI using the primers lco hym (5’–CAA ATC ATA AAG ATA TTG G–3’) and hco2 (5’–CCA GGT AAA ATT AAA ATA TAA...
ACT TC–3') (Schulmeister, 2003), which produce a 676 base pair fragment equivalent to the “barcode” region (Hebert et al., 2003a,b). We analyzed ITS2 using the primers ITS2-F (5'-GGG TCG ATG AAC GCA GC-3') and ITS2-R (5'-ATA TGG TTA AAT TCA GCG GG-3') which anneal to the flanking 28S and 5.8S genes (Navajas et al., 1994). Sequences are deposited in NCBI's GenBank, and accession numbers are listed in Supplementary Table 1.

We sequenced COI for 565 Ophion specimens. A few were not successfully sequenced in both directions, or were otherwise incomplete; these sequences were included with gaps coded as missing data. The 95 Ophion sequences obtained from BOLD and GenBank were sequenced using different primers (see BOLD website for sequencing protocols) and were slightly shorter in length. Alignment was unambiguous, and was confirmed by translating nucleotides to amino acids in Mesquite (Maddison and Maddison, 2011). The data matrix can be found deposited on the Dryad data repository (available at datadryad.org, doi: 10.5061/dryad.hf155).

Since analysis of ITS2 secondary structure (see Section 2.4 and Supplementary Data 2) was sensitive to missing data, we only included sequences that were complete and lacked ambiguous regions; the final dataset included 394 ITS2 sequences. Excluding the flanking 28S and 5.8S regions, the unaligned sequences ranged from 717 to 1053 base pairs in length. A subset of these (380 specimens) were successfully sequenced for both ITS2 and COI (Table 1).

### 2.3. Barcode accumulation curves

All species-delimitation algorithms are dependent on sampling (Bergsten et al., 2012; Lohse, 2009; Meyer and Paulay, 2005; Papadopoulou et al., 2009). We assessed the completeness of sampling by calculating “barcode accumulation curves” of the COI sequences, following the procedure outlined by Smith et al. (2009). Similar to species accumulation curves used in ecological studies, barcode accumulation curves should approach an asymptote as fewer unique haplotypes are added to the specimen pool, indicating that most of the diversity has been sampled. Since the vast majority of specimens were from Canada, we calculated the curve for Canadian material alone (n = 561), as well as a separate curve including all Nearctic specimens (n = 580). We also calculated an accumulation curve for specimens from the province of Alberta, since this is the most densely sampled region (n = 342). Finally, we calculated a barcode accumulation curve including only Palearctic specimens (n = 78). These curves are not based on random samples and do not represent comparable regions. We use them to indicate the pattern of sequence accumulation within the current study, but do not expect them to represent the true diversity of each region.

To calculate the curves, we first constructed a neighbor-joining tree for each dataset in MEGA version 5 (Tamura et al., 2011), using the K2P model and pairwise deletion of missing data, and imported the tree into the program Conserve (Agapow and Crozier, 2008). We excluded all sequences, and then randomly added sequences in sets of 10 (Nearctic, Canada and Alberta datasets) or sets of 5 (UK dataset), and calculated the phylogenetic diversity (PD) of the tree at each successive addition. These results were plotted to assess whether the phylogenetic diversity was approaching an asymptote.

### 2.4. ITS2 secondary structure and CBCs

The complete protocol used to determine the secondary structure of ITS2 is described in Supplementary Data 2. Briefly, we estimated the structure for an arbitrarily chosen sequence using the RNAfold webserver, using the minimum free energy method with default settings (Hofacker, 2003). We imported this sequence and structural into the ITS2 database (Koetschan et al., 2010; Schultz et al., 2006; Selig et al., 2008) and used it as a template for folding the remaining sequences. If any sequences had less than 90% similarity to the existing template, one of the anomalous sequences was directly folded in RNAfold and added as an additional template to the database. We repeated this iterative process until all sequences were folded; the final dataset consisted of eight sequences that were folded directly and 386 sequences that were folded using homology. Sequences and structures were aligned using 4sale, a program designed to perform structural alignments of RNA sequences (Seibel et al., 2006), and a matrix of compensatory base changes (CBC's) was exported. The final alignment (excluding the 28S and 5.8S flanking regions) was 1750 base pairs in length. This alignment was used for all phylogenetic analyses described below, and can be found on the Dryad data repository (datadryad.org, doi: 10.5061/dryad.hf155). We calculated a neighbor-joining tree based on the CBC distance matrix using the program CBCAnalyzer (Wolf et al., 2005).

### 2.5. Species delimitation

We analyzed sequence data using four quantitative methods of species delimitation: generalized mixed Yule coalescent analysis (GMYC) (Fontaneto et al., 2007; Pons et al., 2006); Poisson tree processes model (PTP) (Zhang et al., 2013); automatic barcode gap discovery (ABGD) (Puillandre et al., 2012a); and threshold analysis using the program jMOTU (Jones et al., 2011). Each analysis uses different terminology to refer to the delimited taxa (GMYC = “entities”; PTP = “phylogenetic species”; ABGD = “groups” or “hypothesical species”; jMOTU = “molecular taxonomic units” or “MOTU”), thus acknowledging that they may not represent biologically meaningful species. For clarity and consistency, we are using the term “putative species” in place of each of the above terms, while the term “species” is applied only to taxa that are delimited more robustly by multiple data sources.

#### 2.5.1. GMYC

Models for the construction of the ultrametric trees required for the GMYC analysis were selected using the program PartitionFinder version 1.0.1 (Lanfear et al., 2012), with each gene analyzed as a single partition. For the ITS2 dataset, the selected model was TVMef + G, and COI used HKY + I + G. We obtained the ultrametric trees using BEAST version 1.7.5 (Drummond et al., 2012) on the CIPRES Science Gateway computing cluster (Miller et al., 2010), with ITS2 and COI analyzed separately. Identical haplotypes were removed using the program Alter (Glez-Peña et al., 2010), and outgroups were excluded. The choice of tree prior can impact the results of the phylogenetic analysis (Ceccharelli et al., 2012). For example, the coalescent tree prior is often more appropriate for population-level data and recently diverged taxa, whereas the Yule prior is recommended for species-level data.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Summary of specimens sequenced from each geographic region represented in this study. Palearctic specimens are all from the western Palearctic region, primarily the United Kingdom.</th>
</tr>
</thead>
<tbody>
<tr>
<td>COI only</td>
<td>ITS2 only</td>
</tr>
<tr>
<td>Canada</td>
<td>198</td>
</tr>
<tr>
<td>USA</td>
<td>12</td>
</tr>
<tr>
<td>Costa Rica</td>
<td>2</td>
</tr>
<tr>
<td>Palearctic</td>
<td>68</td>
</tr>
<tr>
<td>Total</td>
<td>280</td>
</tr>
</tbody>
</table>
We therefore calculated trees using a random starting tree, the uncorrelated lognormal relaxed clock model, and two different tree models (coalescent and Yule-process) for a total of four analyses for each marker (ITS2 and COI).

Each analysis was run with either four (COI coalescent, COI Yule, ITS2 coalescent) or five (ITS2 Yule) independent threads, and log files and tree files were combined using LogCombiner. The two COI analyses were run for 40 million generations, with trees sampled every 4000 generations and a burnin of 25%. Combined logfiles were examined in Tracer version 1.5 (Rambaut and Drummond, 2009) to assess convergence; all combined estimated sample sizes (ESS) were >200 with the exception of a single parameter (mean.Rate, ESS = 179) using the Yule prior. The ITS2 analysis with the coalescent prior was run for 80 million generations, sampling every 4000 generations, with a burnin of 25%; all ESS were >200 except for two parameters (mean.Rate, tree.model). Root that were between 100 and 200. Obtaining convergence in the ITS2 analysis with the Yule prior was more challenging. The final analysis included three threads of 160 million generations and two threads of 80 million generations, sampled every 4000 generations. The first 25% of the longer runs, and the first 50% of the shorter runs were removed as burnin. Two parameters failed to show convergence (coefofVariation and covariance, ESS <100); however, the remaining parameters were either >200 or very nearly so. All analyses were resampled in LogCombiner (part of the BEAST package) to give a combined total of approximately 15,000 trees. Maximum clade credibility trees (the tree in each analysis with the highest product of posterior clade probabilities) were produced from the combined tree files in TreeAnnotator (part of the BEAST package), using the mean heights option, and used as input for GMYC analyses.

We conducted GMYC analyses with the SPLITs package (available from http://r-forge.r-project.org/projects/splits/) using the program R (RStudio, 2012; R Core Development Team, 2013). Each tree was analyzed using both single and multiple threshold settings, including all haplotypes (COIall and ITS2all). To facilitate comparison between genetic markers, we also ran the GMYC analyses only including taxa that were represented in both datasets (COIcommon and ITS2common). Since in several instances a single COI haplotype was represented by multiple ITS2 haplotypes, or vice versa, the two analyses did not necessarily have the same number of individuals. We obtained the reduced trees by pruning non-overlapping taxa from ultrametric trees in R, prior to running GMYC analyses.

2.5.2. PTP

Poisson tree processes analyses used maximum-likelihood trees as input. Models for the trees were determined using PartitionFinder version 1.0.1 (Lanfear et al., 2012). For COI, each codon position was included as a potential partition, while ITS2 was not partitioned. Only models that could be implemented in the GARLI web service were included. For COI, the first and second position used GTR + I + G, while the second position used F81 + G. ITS2 was analyzed using SYM + G.

Maximum likelihood trees were calculated using the web service for GARLI 2.1 (Bazinet et al., 2014; available at www. molecularevolution.org), using a neighbor-joining starting tree (calculated in Mega 5.0; Tamura et al., 2011). The GARLI web service uses an adaptive approach wherein additional search replicates are based on the statistical probability that the best tree has been obtained with 95% confidence. For ITS2 a total of 352 replicates were conducted, while the COI analysis reached the maximum number of 1000 replicates. In both cases, there was essentially no difference in topology between the many best-scoring trees. COI trees were rooted with Enicospilus, another genus in the subfamily Ophioninae, whereas ITS2 trees were rooted with O. minutus, since molecular and morphological both suggest that this species is basal within Ophion (Schwarzfeld, unpublished data).

PTP was conducted using the online web service available at http://species.h-its.org/ptp (Zhang et al., 2013). This server conducts both the original maximum likelihood PTP, as well as a newly developed Bayesian version of PTP (bPTP) that adds Bayesian support values to the delimited species. However, our analyses failed to show convergence even with the maximum number of generations (500,000) available on the web server; we have therefore only included the maximum likelihood results in this paper. We conducted PTP on trees including all sequences (COIall, ITS2all), and on trees including only those specimens with sequences available for both genes (COIcommon, ITS2common). Non-overlapping sequences were pruned from the ML trees in R, prior to running PTP.

2.5.3. ABGD

Automatic barcode gap discovery (ABGD) infers clusters of sequences by recursively partitioning the data using a range of prior intraspecific sequence divergences and by calculating a model-based confidence limit for intraspecific divergence at each iteration (Puillandre et al., 2012a). We first calculated distance matrices for COI and ITS2 in MEGA 5.1 using both K2P corrected distances and uncorrected p-distances. Separate matrices were calculated for all specimens (COIall and ITS2all), or only including specimens that were successfully sequenced for both genes (COIcommon and ITS2common). We calculated the number of clusters using the ABGD online tool (Puillandre et al., 2012a; available at http://wwwabi.snv.jussieu.fr/public/abgd/abgdweb.html). Analyses of COI used the default priors (prior intraspecific divergence = 0.001–0.1, divided into 10 partitions), since the program is optimized for use with the COI barcoding gene, based on numerous studies. For ITS2, intraspecific sequence divergence has been less commonly assessed and is highly taxon-dependent. Within Ichneumonidae, Wagener et al. (2006) found the intraspecific sequence divergence of Diadeagma (subfamily Campopleginae) to be 0.2–0.6%. Within the Ophion scutellaris species group (Schwarzfeld and Sperling, 2014), intraspecific variation was similarly very low, ranging from a single haplotype per species to a maximum of 0.1%. We therefore ran the ABGD analysis using prior intraspecific divergences of 0.0001–0.1 (divided into 20 partitions). For each analysis we used both the default barcode relative gap width of 1.5 and a gap width of half the default (0.75) to increase the sensitivity of the analysis (Puillandre et al., 2012a).

2.5.4. Threshold analysis

COI and ITS2 sequences were clustered into molecular taxonomic units (MOTU) using the program jMOTU (Jones et al., 2011). Commonly used thresholds for species delimitation based on COI are 1–3% (Bribiesca-Contreras et al., 2013; Hebert et al., 2003a, b, 2004; Smith et al., 2009, 2013; Stalhut et al., 2013; Strutzenger et al., 2011; Tang et al., 2012). Since computational expense increases very little with more thresholds in the jMOTU analysis, we calculated the number of species using a wider range of thresholds, from a single base pair difference (approximately 0.15%) to a 50 base pair threshold (7.5% of the average sequence length). For the analysis of ITS2, we used the same range of thresholds (1–50 base pairs) as in the COI dataset, which is equivalent to 0.1–5.7% of the average sequence length. For the analysis of ITS2, we used the same range of thresholds (1–50 base pairs) as in the COI dataset, which is equivalent to 0.1–5.7% of the average sequence length. Since the ITS2 sequences contain numerous large indels, we used a range of low identity BLAST filters, from 80 to 97, to attempt to obtain the best alignment (Jones et al., 2010); however, the results were identical in all cases. We therefore ran analyses on both the complete datasets (COIall and ITS2all) and on the common datasets (COIcommon and ITS2common), using the default settings of a minimum 60% overlap and low identity BLAST filter set to 97.
2.5.5. Comparison with morphology

Comparisons of species-delimitation methods provide useful information into the performance of algorithms. However, even if multiple methods converge on the same results, there is no guarantee that the clusters represent species. Morphological analysis of *Ophion* is ongoing, with many species thus far lacking clear morphological characters. Nevertheless, some species are morphologically distinguishable. We selected 11 morphospecies to use as test species to determine whether or at what threshold they were recovered by various analyses. Seven of these species are the Nearctic members of the *O. scutellaris* species group (*O. idoneus*, *O. importunus*, *O. keala*, *O. clave*, *O. aureus*, *O. brevipunctatus*, and *O. dombroskii*) that were resolved using a combination of morphology, wing morphometrics and classical morphometrics (Schwarzfeld and Sperling, 2014). The other species include *O. nigrovarius* Provancher, *O. slossonae* Davis, and two undescribed, morphologically distinct species, sp. A and sp. B. Descriptions and photographs of the diagnostic features of each species are provided in Supplementary Data 3, or in Schwarzfeld and Sperling (2014).

We calculated the average and maximum sequence divergence within each species (excluding those represented by singletons). Distances are uncorrected p-distances, calculated in Mega 5.1 (Tamura et al., 2011), using the pairwise deletion option for gaps and missing data. For the species delimitation analyses, we examined results from the COI all and ITS2 all datasets. We considered an analysis to have successfully recovered a species if a single cluster included all of the sequenced individuals of a morphologically defined species and no additional sequences. For the GMYC and PTP analyses, we reported if a species was successfully recovered (+), split between more than one cluster (Sp), or lumped with other non-conspecific sequences (L). For the ABGD and jMOTU analyses, we listed the range of prior intraspecific divergences or thresholds, respectively, that successfully recovered the species, or recorded if the species was never recovered (−).

2.5.6. Species delimitation of Palearctic *Ophion*

Since the Nearctic fauna is almost entirely undescribed, few well-characterized species are available for assessing the performance of species-delimitation methods. *Ophion* from the UK have been much more thoroughly studied, with a comprehensive morphological revision and workable key to species (Brock, 1982). However, these species have not been assessed with molecular data, and several species present potential taxonomic difficulties (Brock, 1982, G. Broad, pers. comm.). We have therefore analyzed the species limits of 15 of the 16 described British species, based on COI sequence data. We restricted the analysis to COI as we had limited success sequencing these specimens for ITS2. Based on the results of the morphological comparisons described above, we selected a subset of analyses that were most effective for distinguishing Nearctic species. We assessed each Palearctic species using GMYC (coalescent prior, single threshold method), PTP, ABGD (relative gap width: 0.75, prior intraspecific divergence 0.0046), and jMOTU (threshold: 7, 14, and 20 base pairs, or approximately 1%, 2%, and 3% sequence divergence). While the 3% threshold had only limited success in delimiting the Nearctic test species, we included it because this threshold is commonly used in the barcoding literature (e.g. Hebert et al., 2003a; Tang et al., 2012).

3. Results

3.1. Barcode accumulation curves

The three Nearctic barcode accumulation curves for COI (all Nearctic specimens, Canadian only, and Alberta only) showed a similar shape and, as expected, diversity increased with geographic scale of sampling (Fig. 1). No dataset reached an asymptote, although the Alberta sequences may be approaching one. Sequences from the UK were still in the steep initial part of the curve, precluding comparison with the other three datasets. The Nearctic closely follows the Canadian curve, since most specimens were included in both datasets. However, each specimen sequenced from the southern United States (a total of ten specimens from Arizona, Florida, Georgia, Kentucky, and New Mexico) almost certainly represents an additional species, based on both genetic divergence and morphology (Schwarzfeld, unpublished data). This curve therefore greatly underestimates the true diversity for the Nearctic region.

3.2. CBC analysis

There was a maximum of 5 compensatory base changes (CBC’s) between any two specimens. The NJ tree derived from the CBC matrix supported the overall topology of species groups within the genus, as determined through a molecular phylogenetic analysis of *Ophion* (Schwarzfeld, unpublished data). However, there were no CBC’s within any species group except for two specimens that differed from the rest of their group by one (*O. obscuratus* from the U.K., DNA6917) or two (undescribed species from Arizona, DNA5569) CBC’s. Within species groups, CBC’s were thus generally uninformative for distinguishing species.

3.3. GMYC

The results of the GMYC analyses are found in Table 2. Three of the single threshold analyses of the Yule prior trees (ITS2 all, ITS2 common, and COI common) failed to significantly reject the null hypothesis that the sequences cannot be differentiated from a single cluster. Nonetheless, the estimated numbers of clusters (excluding singletons) and putative species (=clusters plus singletons, or "entities") from these analyses were within the range of all other analyses. Across all analyses, the proportion of singletons ranged from 29% to 57% of the total number of GMYC species, with an average of 40%.

In general, the number of clusters was fairly robust with respect to both tree prior (Yule (Y) or coalescent (C)) and GMYC method (single (S) or multiple (M)). For the COI all data, the number of clusters varied from 72 to 81, while in the smaller ITS2 all dataset three of the four analyses resulted in 52 or 53 clusters. The YM analysis of the ITS2 all data, however, only recovered 35 clusters. When only common taxa were included, the number of clusters was also quite robust between the two genes, with all analyses recovering 45–50 clusters.
The total number of putative species was more variable. For the COI all dataset, the number was fairly similar between three of the analyses (114–124); however, the CM analysis recovered 150 putative species. For the ITS2 all analyses, the number of putative species varied from 51 (YM) to 95 (CM). For the "common" datasets, there was a minimum of 69 putative species (ITS2 YS) and a maximum of 107 (COI CM). All of the COI analyses recovered more putative species than did the corresponding ITS2 analyses. Within COI common, the YM and CM analyses recovered far more putative species (103 and 107, respectively) than did the corresponding single threshold analyses (86 and 80).

### 3.4. PTP

Analysis of COI all gave a total of 89 putative species, while ITS2 all analysis resulted in 54 putative species. It was fairly congruent with GMYC where lineages were relatively deep (Fig. 2A), but more conservative with respect to splitting taxa when divergences were shallow (Fig. 2B). When only "common" taxa were included, several more putative species were delimited in the COI analysis (COI common: 84 putative species) compared to ITS2 (61 putative species). Interestingly, the "common" datasets tended to split taxa in comparison to the "all" datasets, to the point that more putative species were delimited in the ITS2 common analysis compared to ITS2 all.

### 3.5. ABGD

For the COI datasets, ABGD delimited the same number of putative species regardless of which distance matrix was used as the input data (K2P or uncorrected p-distance). In the ITS2 analyses, a few intermediate partitions had a maximum of three additional putative species with the uncorrected distance matrix, compared with the K2P distance matrix. Since the results were so similar, we are only reporting the results of the uncorrected analysis.

The relative gap width had a strong influence on the intermediate partitions of both the COI and ITS2 analyses (Fig. 3). For example, at a prior intraspecific divergence of 0.0017 (COI) or 0.0013 (ITS2), the more sensitive analysis (gap width = 0.75) recovered over twice as many putative species as the default gap width (=1.5).

The number of putative species recovered by each COI analysis dropped precipitously between the first and second partition (COI priors = 0.0010 and 0.0017, respectively; Fig. 3a). Over the remaining partitions, the pattern was analysis- and dataset-dependent, though by a prior of 0.036 all analyses were recovering a single cluster. The COI all analysis using the default gap width had a clear plateau in the number of species recovered, ranging from 66 to 82. However, the pattern was less clear in the other analyses. The more sensitive analyses (gap width = 0.75) had a range of 52–168 putative species (COI all) or 23–120 species (COI common) over the intermediate priors, with the second and third partition recovering the same number of putative species. The COI common analysis with the default gap width (1.5) had only a single intermediate partition, with 48 putative species.

The ITS2 all and ITS2 common ABGD analyses gave identical results, except that 0–3 fewer putative species were delimited in each partition for ITS2 common. For clarity, we have therefore only shown the ITS2 all data in Fig. 3b. Using any prior intraspecific divergences at or below 0.001 resulted in the same number of putative species (COI all: 120; COI common: 117). With priors above 0.001, the number of putative species dropped rapidly. With a prior of 0.0013, 27 and 59 putative species were recovered by the two ITS2 analyses (gap width 1.5 and 0.75, respectively) and with a prior of 0.0078 or greater, fewer than 20 putative species were recovered by all analyses.

### 3.6. Threshold analysis

The number of putative species delimited for both COI and ITS2 initially dropped steeply as the threshold increased, eventually flattening out (Fig. 4). The curve was more gradual for COI, not reaching a clear plateau until over 5% sequence divergence, whereas for ITS2, the curve flattened out by approximately 1% divergence. A total of 128, 90, and 47 putative species were delimited by the analysis of COI all at thresholds of approximately 1%, 2%, and 3%, respectively (7, 13, and 20 base pairs). At these thresholds, the COI common dataset delimited 74, 43, and 20 putative species. Comparing this to the ITS2 common dataset, the most similar number of putative species were found at thresholds of 0.3%, 0.8%, and 2.6–3% (3, 7, and 23–26 base pairs).

### 3.7. Test species

The mean percent sequence divergence within each test species ranged from 0.04% to 0.6% divergence for COI. The highest

---

Table 2

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Tree prior</th>
<th>Method</th>
<th>Clu</th>
<th>Spp</th>
<th>CI</th>
<th>L-Null</th>
<th>L-GMYC</th>
<th>LRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>COI all</td>
<td>Coalescent</td>
<td>Single</td>
<td>72</td>
<td>114</td>
<td>100–137</td>
<td>3119.398</td>
<td>3135.499</td>
<td>&lt;0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multiple</td>
<td>83</td>
<td>150</td>
<td>129–167</td>
<td>3119.398</td>
<td>3145.07</td>
<td>&lt;0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Single</td>
<td>81</td>
<td>121</td>
<td>117–134</td>
<td>3057.645</td>
<td>3064.79</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multiple</td>
<td>79</td>
<td>124</td>
<td>88–125</td>
<td>3057.645</td>
<td>3073.966</td>
<td>0.000</td>
</tr>
<tr>
<td>ITS2 all</td>
<td>Coalescent</td>
<td>Single</td>
<td>53</td>
<td>85</td>
<td>85–104</td>
<td>2326.429</td>
<td>2334.213</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multiple</td>
<td>52</td>
<td>95</td>
<td>93–96</td>
<td>2326.429</td>
<td>2339.508</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Single</td>
<td>53</td>
<td>75</td>
<td>62–81</td>
<td>2162.679</td>
<td>2166.034</td>
<td>0.081a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multiple</td>
<td>35</td>
<td>51</td>
<td>47–104</td>
<td>2162.679</td>
<td>2173.709</td>
<td>0.001</td>
</tr>
<tr>
<td>COI common</td>
<td>Coalescent</td>
<td>Single</td>
<td>45</td>
<td>80</td>
<td>67–93</td>
<td>1728.258</td>
<td>1739.615</td>
<td>&lt;0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multiple</td>
<td>49</td>
<td>107</td>
<td>103–107</td>
<td>1728.258</td>
<td>1744.155</td>
<td>&lt;0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Single</td>
<td>50</td>
<td>86</td>
<td>2–106</td>
<td>1695.617</td>
<td>1698.06</td>
<td>0.180a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multiple</td>
<td>44</td>
<td>103</td>
<td>72–116</td>
<td>1695.617</td>
<td>1704.424</td>
<td>0.024</td>
</tr>
<tr>
<td>ITS2 common</td>
<td>Coalescent</td>
<td>Single</td>
<td>44</td>
<td>76</td>
<td>53–94</td>
<td>2135.929</td>
<td>2143.237</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multiple</td>
<td>49</td>
<td>78</td>
<td>78–92</td>
<td>2135.929</td>
<td>2147.648</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Single</td>
<td>48</td>
<td>69</td>
<td>59–77</td>
<td>1990.495</td>
<td>1994.305</td>
<td>0.054a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multiple</td>
<td>48</td>
<td>73</td>
<td>39–99</td>
<td>1990.495</td>
<td>2001.224</td>
<td>0.0181</td>
</tr>
</tbody>
</table>

*Not significant.*
Fig. 2. Demonstration of putative species limits as obtained through different quantitative delimitation methods. The phylogenetic tree is the coalescent tree of COI obtained through BEAST analysis; it is being used for representation, but is not intended to be a definitive phylogenetic tree for the genus. Section A of the tree demonstrates relatively good congruence between methods, while section B demonstrates widespread incongruence. Red = GMYC: c–s = coalescent prior, single threshold; c–m = coalescent prior, multiple threshold; y–s = Yule prior, single threshold; y–m = Yule prior, multiple threshold; Green = PTP; Orange = jMOTU: 1 = 1%, 2 = 2%, 3 = 3%; Blue = ABGD: a = 1.5 gap width, 2nd partition; b = 0.75 gap width, 2nd partition; c = 1.5 gap width, 4th partition; d = 0.75 gap width, 4th partition. Each bar represents a putative species. Black bars indicate the associated specimen(s) were not included within the putative species (i.e. taxa were not monophyletic with respect to this tree). Numbers are used to indicate when two separate bars belong to the same species (non-monophyletic on this tree). Identical haplotypes were removed for this phylogenetic analysis; therefore, some apparent singletons may be represented by multiple specimens in the complete dataset.
maximum divergence within any species was found in Sp. B, with 0.9% divergence. ITS2 intraspecific divergence was almost non-existent, with only Sp. B showing any divergence (mean 0.1%, maximum 0.3%) (Table 3).

GMYC analysis of COI was generally more successful at correctly delimiting species than that of ITS2 (Table 4). Within the COI analysis, the coalescent or Yule tree priors were equally successful at recovering the test species. The single threshold method was most successful, correctly delimiting all species except O. brevipunctatus and O. dombroskii, which were recovered as a single cluster. The multiple-threshold method also lumped these two species, and in addition failed to recover O. keala and Sp. A. The CS, CM, and YS analyses of the ITS2 data each delimited 5 of the 10 species but which species were successfully delimited varied between the analyses. The YM analysis of ITS2 was the least successful GMYC analysis, recovering only two species. O. nigrovarius was the only species to be recovered by all GMYC analyses and both markers.

Similarly, species were more successfully delimited in the PTP analysis of COI, compared to ITS2. However PTP in general was less successful than GMYC, with the COI analysis correctly delimiting four species out of eleven, while in the ITS2 analysis only a single species was successfully delimited (Table 4).

Sp. B was never recovered by ABGD, being either split or lumped, depending on the parameters used (Table 5). All other species were recovered by at least one ABGD analysis. Again, the analyses of COI were generally more successful than the ITS2 analyses, with most species being successfully delimited over a range of partitions. The more sensitive analysis of COI (gap width = 0.75) was most successful, with all species except for Sp. B being correctly delimited at least one, and usually several, partitions. Within this analysis, the prior intraspecific divergence of 0.0046 was the only partition that correctly delimited all species (excluding Sp. B). ABGD of the ITS2 dataset recovered 7 of the 10 species in all partitions with prior intraspecific divergences < 0.001%; three of these species were also recovered in additional partitions. The remaining three species were never successfully delimited.

All of the test species were successfully delimited by jMOTU with the exception of Sp. B in the analysis of COI and O. clave in the analysis of ITS2 (Table 5). In both cases, these species were split at low thresholds, and lumped with other species at higher thresholds, but never delimited successfully. In the COI analysis, all remaining species were delimited at a threshold of 1% sequence divergence, and all except O. brevipunctatus and O. dombroskii at 2% divergence. Only three species (O. nigrovarius, O. slossonae, and O. keala) were delimited at 3% divergence or higher. For the ITS2 analysis, most species were delimited at very low divergences, with only O. nigrovarius and Sp. A being successfully delimited at thresholds of more than 1% divergence (9 base pairs).

3.8. Species delimitation of Palearctic Ophion

Compared to the Nearctic test species, sequence divergence within several of the Palearctic “species” was high (Table 6). Nine species had maximum intraspecific divergences of over 2%. Two species (O. pteridis and O. perkinsi) were represented by a single haplotype, and sequence divergence within O. ventricosus and O. ocellaris were also very low (Table 6).
groups, in several instances it lumped these putative species with
the jMOTU analyses often split species into a similar number of
species within a few clades (Table 7). In comparison, while
tation of Palearctic species, with only slight differences in the num-
ber of species within a few clades (Table 7). In comparison, while

<table>
<thead>
<tr>
<th>Species</th>
<th>COI</th>
<th>ITS2</th>
<th>COI</th>
<th>ITS2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CS</td>
<td>CM</td>
<td>YS</td>
<td>YM</td>
</tr>
<tr>
<td>clave</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aurues</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>brevipunctatus</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dombroskii</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>keula</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>importunus</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>idoneus</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nigrovarius</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>slossonae</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sp. A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sp. B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sp. A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sp. B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>brevicornis</td>
<td>3</td>
<td>2</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>costatus</td>
<td>13</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>crassicornis</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>forticornis</td>
<td>1</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>idoneus</td>
<td>14</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>longigena</td>
<td>8</td>
<td>2</td>
<td>+</td>
<td>3</td>
</tr>
<tr>
<td>luteus</td>
<td>14</td>
<td>2</td>
<td>+</td>
<td>2</td>
</tr>
<tr>
<td>minutus</td>
<td>8</td>
<td>2</td>
<td>+</td>
<td>3</td>
</tr>
<tr>
<td>mocsaryi</td>
<td>9</td>
<td>3</td>
<td>+</td>
<td>3</td>
</tr>
<tr>
<td>obscuratus</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>ocellaris</td>
<td>3</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>parvulus</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>perkinsi</td>
<td>2</td>
<td>+</td>
<td>2</td>
<td>+</td>
</tr>
<tr>
<td>pteridis</td>
<td>3</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>scutellarius</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>ventricosus</td>
<td>4</td>
<td>+</td>
<td>+</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 4: Summary of congruence between generalized mixed-Yule coalescence analyses (GMYC) and Poisson tree processes (PTP) with eleven species/morphospecies of Nearctic Ophion. C = coalescent tree prior; Y = Yule tree prior; S = single threshold model; M = multiple threshold model; + = species successfully delimited; L = species lumped with other sequences; Sp = species split into >1 groups.

Table 5: Summary of congruence between automatic barcode gap discovery (ABGD) and threshold analyses with eleven species/morphospecies of Nearctic Ophion. ABGD used two gap widths: 1.5 and 0.75; numbers represent the range of prior intraspecific percent divergence that successfully delimited each species. jMOTU = threshold analyses using program jMOTU; numbers indicate the range of thresholds; as a percentage of the mean sequence length, that successfully delimited each species; for both analyses, – indicates the species was never delimited.

Table 6: Mean and maximum percent sequence divergence of COI within each Palearctic species that was represented by more than one individual. Divergences are uncorrected p-distances (%), with gaps and missing data treated by pairwise deletion.

Table 7: Number of putative species delimited for each morphologically identified species of Ophion, based on the COI barcode region. All specimens are from the United Kingdom, with the exception of a single specimen of O. obscuratus from Spain and one specimen of O. forticornis from France. Analyses that successfully delimited a species are identified with +. The GMYC analysis used the single threshold method on an ultrametric tree produced in BEAST, with a relaxed lognormal clock and coalescent tree prior. PTP was conducted on a maximum likelihood tree obtained with GARLI. The ABGD analysis used a relative gap width of 0.75 and a prior intraspecific divergence of 0.0046. The jMOTU analysis used thresholds of 7, 13 and 20 base pairs, or approximately 1%, 2%, and 3% sequence divergence.

ABGD, GMYC, and PTP were broadly congruent in their delimitation of Palearctic species, with only slight differences in the number of species within a few clades (Table 7). In comparison, while the jMOTU analyses often split species into a similar number of groups, in several instances it lumped these putative species with other sequences. In some cases jMOTU analyses combined highly divergent taxa. For example, at thresholds of 2% and 3% divergence, it combined O. pteridis and O. luteus, along with several Nearctic species, into one large group, even though they differ by approximately 6% sequence divergence.

Of the 15 Palearctic Ophion species that were analyzed, three (O. forticornis, O. longigena, O. ocellaris) were delimited by all analyses, although two of these species (O. forticornis and O. longigena) were represented by singletons. O. ventricosus was recovered as a single species by all except the 1% jMOTU analysis, and O. perkinsi by all except PTP (oddly, since it was represented by a single haplotype). Also represented by a single haplotype, O. pteridis were never split into multiple species; however, according to the 2% and 3% jMOTU analyses, it was lumped with other sequences. All other species were split to various extents by the different analyses.

ABGD, GMYC, and PTP were broadly congruent in their delimitation of Palearctic species, with only slight differences in the number of species within a few clades (Table 7). In comparison, while the jMOTU analyses often split species into a similar number of groups, in several instances it lumped these putative species with other sequences. In some cases jMOTU analyses combined highly divergent taxa. For example, at thresholds of 2% and 3% divergence, it combined O. pteridis and O. luteus, along with several Nearctic species, into one large group, even though they differ by approximately 6% sequence divergence.

Of the 15 Palearctic Ophion species that were analyzed, three (O. forticornis, O. longigena, O. ocellaris) were delimited by all analyses, although two of these species (O. forticornis and O. longigena) were represented by singletons. O. ventricosus was recovered as a single species by all except the 1% jMOTU analysis, and O. perkinsi by all except PTP (oddly, since it was represented by a single haplotype). Also represented by a single haplotype, O. pteridis were never split into multiple species; however, according to the 2% and 3% jMOTU analyses, it was lumped with other sequences. All other species were split to various extents by the different analyses.

ABGD, GMYC, and PTP were broadly congruent in their delimitation of Palearctic species, with only slight differences in the number of species within a few clades (Table 7). In comparison, while the jMOTU analyses often split species into a similar number of groups, in several instances it lumped these putative species with other sequences. In some cases jMOTU analyses combined highly divergent taxa. For example, at thresholds of 2% and 3% divergence, it combined O. pteridis and O. luteus, along with several Nearctic species, into one large group, even though they differ by approximately 6% sequence divergence.

Of the 15 Palearctic Ophion species that were analyzed, three (O. forticornis, O. longigena, O. ocellaris) were delimited by all analyses, although two of these species (O. forticornis and O. longigena) were represented by singletons. O. ventricosus was recovered as a single species by all except the 1% jMOTU analysis, and O. perkinsi by all except PTP (oddly, since it was represented by a single haplotype). Also represented by a single haplotype, O. pteridis were never split into multiple species; however, according to the 2% and 3% jMOTU analyses, it was lumped with other sequences. All other species were split to various extents by the different analyses.

ABGD, GMYC, and PTP were broadly congruent in their delimitation of Palearctic species, with only slight differences in the number of species within a few clades (Table 7). In comparison, while the jMOTU analyses often split species into a similar number of groups, in several instances it lumped these putative species with other sequences. In some cases jMOTU analyses combined highly divergent taxa. For example, at thresholds of 2% and 3% divergence, it combined O. pteridis and O. luteus, along with several Nearctic species, into one large group, even though they differ by approximately 6% sequence divergence.

Of the 15 Palearctic Ophion species that were analyzed, three (O. forticornis, O. longigena, O. ocellaris) were delimited by all analyses, although two of these species (O. forticornis and O. longigena) were represented by singletons. O. ventricosus was recovered as a single species by all except the 1% jMOTU analysis, and O. perkinsi by all except PTP (oddly, since it was represented by a single haplotype). Also represented by a single haplotype, O. pteridis were never split into multiple species; however, according to the 2% and 3% jMOTU analyses, it was lumped with other sequences. All other species were split to various extents by the different analyses.
4. Discussion

4.1. Sampling

All quantitative species-delimitation methods can be misled if species are not adequately sampled, a common issue in diverse, understudied taxa (Lim et al., 2012; Lohse, 2009; Papadopoulou et al., 2009). In particular, undersampling can lead to either underestimation of species numbers (since some species are not sampled) or overestimation (since some of the diversity within species may not be sampled, resulting in apparent genetic structuring) (Papadopoulou et al., 2009). However, determining what level of sampling is considered “adequate” is not a trivial task. Species ranges, geographical structuring, effective population sizes, intraspecific genetic variation, habitat niches (possibly including host-races in the case of parasitoids) all play a role in the observed data (Bergsten et al., 2012; Hamilton et al., 2014). In theory, even sampling across the genetic and geographic ranges of each species would improve delimitation methods; however in practice this is likely impossible to achieve. For example, Bergsten et al. (2012) found up to 70 individuals per species of diving beetle were required to sample 95% of intraspecific variation.

In our study, none of the barcode accumulation curves reached an asymptote, a common finding among diverse taxa such as parasitoids (Fernández-Triana et al., 2011; Smith et al., 2009; Zaldívar-Riverón et al., 2010). This study is almost entirely restricted to specimens from Canada and the northern United States. While the southern Nearctic region is almost entirely unexamined, the more extensively studied Canadian fauna is also insufficiently sampled. The most densely sampled region is the province of Alberta, which provided 52% of the sequences for this study; the curve representing these sequences appears to be approaching an asymptote, but even here, more sampling is needed to fully assess the genetic diversity of these species. Included samples were collected haphazardly, in many cases as bycatch from moth collectors. Future studies should therefore place a high priority on sampling throughout a wider range of habitats and geographic regions.

A high proportion of singleton species were recovered in all analyses, another typical result for diverse, understudied taxa (Ceccherelli et al., 2012; Lim et al., 2012; Monaghan et al., 2009). The robustness of species-delimitation algorithms with regard to undersampled species is an important consideration as singleton species are abundant in taxonomic and biodiversity studies (Lim et al., 2012). The GMYC method has been found to be accurate even with a high proportion of singleton species (Ceccherelli et al., 2012; Monaghan et al., 2009; Talavera et al., 2013). In contrast, for accurate results with ABGD, it is recommended that 3–5 sequences should be included per species (Puillandre et al., 2012a). However, without a priori knowledge of the group being studied, this may be difficult to achieve, and at least in our study, the singletons did not appear to negatively affect the results. PTP has been found to overestimate diversity when sampling is uneven between species (Zhang et al., 2013), yet despite the uneven sampling in this study, PTP was more prone to underestimating the number of species, compared to GMYC. However, this pattern may be indicated by the increased splitting of putative species when sequences were removed in the “common” analyses. Further studies of the effect of taxon sampling with regards to PTP are needed.

Increased sampling does not always lead to more accurate species delimitation, particularly in threshold-based analyses. An apparently distinct barcoding gap often disappears with increased sampling across the range of genetic variation of a given taxon (Bergsten et al., 2012; Meyer and Paulay, 2005; Moritz and Cicero, 2004; Wiemers and Fiedler, 2007). In this study, at the commonly used 2% threshold, the single-linkage clustering algorithm used by jMOTU clustered several distantly-related taxa. For example, O. luteus, O. pteridis, and a number of Nearctic species, were grouped into a single MOTU, as there were apparently enough intermediate sequences to act as a bridge between them. However, a trial analysis with the dataset restricted to O. luteus and O. pteridis retained them as separate MOTUs at thresholds of up to 6% divergence.

Hamilton et al. (2014) recommend incorporating prior knowledge of natural history, genetic variation, and morphology in order to design an effective sampling strategy for accurate species delimitation. This may be difficult to achieve with the first attempt at delimiting species in an almost entirely unexplored genus, such as Ophion. However, as knowledge of the taxon accumulates, it should be possible to gradually “fill in the holes”, and thus refine and improve confidence in the delimitation results.

4.2. Comparison of ITS2 and COI as markers for species delimitation of Ophion

ITS2 in Ophion has a very different pattern of divergence compared to COI, with high divergence between species groups, but very low intra- and interspecific divergence among closely related species (Schwarzfeld and Sperling, 2014; Schwarzfeld, unpublished data). Nevertheless, the branching pattern was largely congruent with COI, and at least according to GMYC, a similar number of species were delimited.

Due to its length, the presence of microsatellite regions, and the high number of indels, ITS2 was more logistically challenging to sequence and align for Ophion. However, both ITS2 and COI have advantages and disadvantages with regards to species delimitation. In this study, COI successfully identified more test species in GMYC and PTP, and the choice of priors or thresholds for ABGD and jMOTU is more straightforward, due to the large number of studies focusing on this gene. In some studies, however, ITS2 has been found to be more effective than COI at distinguishing closely related species (e.g. Wagener et al., 2006). As well, certain indels of ITS2 can be a species-specific character, and have been used to successfully distinguish species of parasitic wasps in several studies (Klopfstein, 2014; Quicke et al., 2006; Van Veen et al., 2003). While beyond the scope of the current study, more detailed examination of indels may resolve some of the ambiguously delimited species of Ophion.

Despite the apparent advantages of COI in this study, relying on this single gene may provide misleading results due to incomplete lineage sorting, introgression, bacterial endosymbionts, or nuclear mitochondrial pseudogenes (nurms) (Dupuis et al., 2012; Funk and Omland, 2003; Raychoudhury et al., 2010; Schmidt and Sperling, 2008; Song et al., 2008). In at least one species in this study, an apparent numt (Gellissen et al., 1983; Lopez et al., 1994) was amplified, as there is a homologous 2 base pair deletion in each sequence which would result in a shift in the codon reading frame of the COI gene. Since the GMYC analysis of ITS2 also recovered the same species, it appears this numt may be taxonomically informative. However not all numts are as readily recognizable, and can be amplified by chance in taxa that are not closely related. It is therefore critical to include an independent molecular marker to verify the results obtained with COI. ITS2 is an important tool for delimiting species in this genus, though the investigation of other relatively rapidly evolving nuclear genes (e.g. CAD) (Sharanowski et al., 2011; Wild and Maddison, 2008) may provide additional and potentially more tractable species-level genetic markers.

4.3. CBC analysis

The potential for a single compensatory base change (CBC) to unambiguously identify species as being reproductively isolated has been a promising “magic bullet” for species delimitation
(Coleman, 2009; Müller et al., 2007; Wolf et al., 2013). The initial studies focused on plants and fungi, but did not assess the utility of CBC’s in animal taxa. The first test of CBC’s for species delimitation of animals was provided by Ruhl et al. (2010), where they found it to effectively delimit two morphologically indistinguishable species of flea beetle (Chrysomelidae: *Altica*). However, studies of blue butterflies (Lycaenidae; Wiemers et al., 2009), mosquitoes (Alquezar et al., 2010), and budworm parasitoids (Smith et al., 2011) found CBC’s were too rare to distinguish closely-related species. The almost complete lack of CBC’s between closely related *Ophion* species provides further evidence that for many insects, CBC’s are apparently too scarce to be of practical use in distinguishing closely-related species. They are, however, potentially useful in defining higher taxonomic groups.

### 4.4. Parameter selection and species delimitation

Two main parameters that must be selected for GMYC analyses are the priors used to construct the phylogenetic tree, and the choice of single- or multiple-threshold methods. While the multiple-threshold method of GMYC was developed to take into account different branching patterns and rates of evolution across a tree (Monaghan et al., 2009; Papadopoulou et al., 2009), evidence suggests it is consistently less accurate than the single-threshold method (Esselstyn et al., 2012; Fujisawa and Barraclough, 2013). This study supports this finding with fewer test species being correctly delimited with this method in both the analyses of COI (both tree priors) and ITS2 (Yule prior). While the multiple-threshold method did not always recover higher numbers of species than did the corresponding single analysis, all of the highest estimates were obtained using this method, further supporting the conclusion of Esselstyn et al. (2012) that the multiple-threshold method overestimates the number of species.

Yule priors are generally considered more appropriate for analyses of speciation, whereas coalescent priors are more applicable to intraspecific population level data (Drummond et al., 2006). In a GMYC analysis of a diverse genus of Braconidae, Ceccarelli et al. (2012) found the accuracy of each prior depended on the gene being analyzed, with the coalescent prior being more consistent with morphology for COI data. However, this result may be taxon-specific; in the current study, both the estimated number of species and the accuracy of the analyses were largely unaffected by which tree prior was used to construct the phylogenetic tree. The birth–death prior is another alternative that may be appropriate for species diversification (Gernhard, 2008; Nee et al., 1994), though it was not tested in the current study.

While fewer parameters must be selected for PTP, the analysis does depend on the accuracy of the input phylogenetic tree. We only tested a single ML tree, and therefore cannot determine if the use of different models or tree-building algorithms would affect the results. Tang et al. (2014) found that PTP was quite robust with respect to the phylogenetic methods used; however, the tendency of taxa to be more split when fewer specimens were included again indicates the importance of taxon sampling for phylogenetic reconstruction.

The results of the ABGD analyses were highly dependent on the gap width selected. In most recent studies that have employed this method, the default gap width of 1.5 is either explicitly used (Hendrixson et al., 2013; Kekkonen and Hebert, 2014; Tang et al., 2012), or implicitly used (Crawford et al., 2012; Hamilton et al., 2014; Puckridge et al., 2013). However, Puillandre et al. (2012b) increased this setting to 10 in order to decrease the sensitivity of their analysis to small gaps, while Weigand et al. (2013) used a range of more sensitive gap widths (0.01–0.9). In this study, reducing the gap width doubled the number of delimited species in several partitions for both COI and ITS2. As well, there was no clear plateau in the estimated number of species with the more sensitive gap width, making the choice of prior a subjective but essential task if this analytical method is to be used to assess species diversity. While both gap widths recovered most test species, in the COI analysis, more species were recovered over a wider range of priors using the more sensitive gap width. The sensitivity of the ABGD analysis to different gap widths is most apparent in one species-group of apparently weakly diverged species (Schwarzfeld, unpublished data). For example, in the analysis of COI using an intraspecific prior of 0.0017, the default analysis estimated 15 species within this group, whereas the more sensitive analysis recovered 81 species (Fig. 2b shows part of this group). The one test species within this species group (Sp. B) was not accurately delimited in either case, being split with the more sensitive analysis and lumped with the default analysis. The sensitivity of the analysis to the gap width may only be an issue within species complexes or recent radiations, which can confound any species-delimitation method (Monaghan et al., 2006; Reid and Carstens, 2012). Nonetheless, the impact of gap width should be considered when using this method, since the choice of this parameter is largely arbitrary.

While gap width also affected the intermediate partitions of the ITS2 analysis this had little impact on the overall results, as only three test species were recovered with any intraspecific divergence above 0.001. This supports the results seen in the *O. scutellaris* group (Schwarzfeld and Sperling, 2014) where all species had a maximum of 0.1% intraspecific divergence.

Based on thresholds of 1–3%, there was almost a threefold difference in the number of species estimated by *jMOTU* from 47 species at 3% divergence to 128 species at 1%. While 2% is one of the most commonly used thresholds in DNA barcoding studies (e.g. Bribiesca-Contreras et al., 2013; Fernandez-Flores et al., 2013; Smith et al., 2013; Strutzenberger et al., 2011), other thresholds are also used, such as 1.6% (Smith et al., 2009), 2.3% (Young et al., 2012), 3% (Hendrich et al., 2010; Tang et al., 2012), or 3.2% (Weigand et al., 2013). Although the choice of threshold is sometimes based on empirical studies (e.g. Hebert et al., 2003a; Weigand et al., 2011), there is no reason to believe the same thresholds will apply in different taxa; even within a taxon, there can be highly variable patterns of intraspecif and interspecific divergence (e.g. Bergmann et al., 2013).

Despite this range, with a standard 2% threshold the estimated number of species (90) was within the range of the GMYC and the intermediate ABGD analyses, and most test species were recovered. However, due to the “greedy algorithm” used in the single-linkage clustering employed by this method (Jones et al., 2011), the delimitation of species boundaries sometimes differed considerably from the other two methods (Fig. 2).

### 4.5. Comparison of methods

We concluded that GMYC and PTP offer some distinct advantages over the other two methods of species delimitation. The largest advantage is that these methods incorporate evolutionary theory and are therefore much less reliant on the application of arbitrary thresholds. Because of this, they are also more tractable with genetic markers such as ITS2 that have less well-characterized intra- and interspecific patterns of diversity. With appropriate tree-building methodologies, these methods can easily be applied to any genetic marker. In comparison, choosing an appropriate threshold or range of thresholds for a marker other than COI is particularly arbitrary, since no “standard” divergence has been accepted for species delimitation.

Nonetheless, caution must be employed when using either of these methods. One disadvantage of single-locus tree-based
methods is that all species, by definition, must be monophyletic, despite evidence that a high proportion of species are not monophyletic with respect to their gene trees (Funk and Omland, 2003). GMYC relies on the priors and parameters used to construct the ultrametric tree (Cecarelli et al., 2012), and the construction of the trees can be computationally demanding. GMYC may also overestimate species diversity compared to other methods (Kekkonen and Hebert, 2014; Hamilton et al., 2014; Miralles and Vences, 2013; Paz and Crawford, 2012; Talavera et al., 2013; Weigand et al., 2013). PTP is far less computationally demanding than the GMYC and since an ultrametric tree is not required, relies on fewer tree-building assumptions. It has been found to be as accurate, if not more so, than GMYC when tested against well-defined species (Dumas et al., 2015; Modica et al., 2014; Tang et al., 2014; Zhang et al., 2013). In our study, it delimited considerably fewer putative species than the GMYC analyses, possibly an advantage due to the hypothesized oversplitting by GMYC. However, since it lumped several of the test species that were correctly delimited by GMYC, its success likely depends to a large extent on the phylogenetic patterns within the taxon in question.

Both jMOTU and ABGD provide the flexibility of examining a range of thresholds or priors, respectively, which can be informative in understanding the effects of the algorithms and the diversity within the group. However, the majority of barcoding studies use a single threshold to estimate species diversity (e.g. Bribiesca-Contreras et al., 2013; Fernandez-Flores et al., 2013; Smith et al., 2013; Strutzenberger et al., 2011; but see Laetsch et al., 2012). A range of priors is more often examined with ABGD (Paz and Crawford, 2012; Puckridge et al., 2013); however, the impact of the relative gap width is rarely considered, even though it can have a significant impact on the results obtained. A further disadvantage of threshold analysis as implemented in jMOTU is the grouping of distantly related species among taxa that lack a clear barcoding gap. Despite these issues, distance-based species-delimitation methods do not rest on the assumption of reciprocal monophyly between species. It is therefore recommended that at least one distance-based and one tree-based method be used in any species-delimitation study (Fontaneto et al., 2015).

All of the above methods are single-locus methods. While using single-locus methods on multiple loci independently allows qualitative comparisons and reduces the potential biases from relying on a single molecular marker, several multi-locus methods have been developed that can better account for some of the issues described above (e.g. gene-tree discordance, non-monophyletic species, recently-diverged species with gene flow, etc.) (Dupuis et al., 2012; Fontaneto et al., 2015). Examples of multi-locus methods include: Bayesian phylogenetics and phylogeography (BP&P; Yang and Rannala, 2010; Rannala and Yang, 2013); SpedeSTEM (Ence and Carstens, 2011); O’Meara’s heuristic model (O’Meara, 2010); and Bayes Factor Delimitation (BFD; Grummer et al., 2014). BP&P is one of the most widely-used methods; however, until recently it required a user-supplied guide tree, thus being less effective for unguided species-delimitation in entirely unknown taxa. Recently, methods have been developed that do not require previous knowledge of the included species, such as an unguided version of BP&P (Yang and Rannala, 2014) and DISSECT (Jones et al., 2015). Most multi-locus methods are computationally-expensive, and, depending on the study system, are arguably “overkill” (Collins and Cruickshank, 2014). Nevertheless, as sequencing becomes increasingly affordable, multi-locus methods will no doubt become the norm. Future studies of Ophion, and of species-delimitation studies in general, should include multi-locus methods.

4.6. How many Ophion species are there?

Based on COI (since many species were not successfully sequenced for ITS2), there was a wide range of estimates for the total number of ITS2 species present in this study. Considering all plausible results of the quantitative species-delimitation analyses (i.e. including all GMYC analyses; PTP; ABGD with gap widths of 1.5 and 0.75 and priors of 0.0017–0.0077 and 0.0017–0.0129 intraspecific divergence, respectively; and jMOTU thresholds of 1–3% divergence), estimates ranged from 47 to 168 species of Ophion represented in this study. Excluding the more extreme estimates (i.e. including only single threshold GMYC analyses; PTP; ABGD analyses with a prior of 0.0046; and jMOTU results with a threshold of 2%) resulted in estimates of 89–121 species (64–97 of which are from the Nearctic region).

The wide range of estimates emphasizes that all methods of molecular species delimitation provide only “primary species hypotheses” (Puillandre et al., 2012b), which need to be assessed using additional methods. As a first step, these methods can be a helpful tool to delimit hypothetical species; however, for comparative studies relying on species numbers (e.g. for conservation applications or for analysis of patterns of biodiversity), it is vital that researchers also consider the parameters and potential shortcomings of the delimitation method(s) used to obtain the estimates.

Our finding that most described British species were split may be due to the failure of delimitation methods for successfully circumscribing true biological species, or alternatively may indicate the presence of cryptic or previously unrecognized species. In several instances, putative species were not recovered as a monophyletic clade by phylogenetic analyses, indicating that multiple, previously unrecognized species are almost certainly present. As well, the steep slope of the barcode accumulation curve for these specimens suggests that genetic sampling is far from complete.

The use of a variety of analytical methods and parameters can aid in planning subsequent research to assess the hypothetical species obtained with these methods (Carstens et al., 2013). For example, groups of sequences that are consistently delimited with a variety of quantitative methods are good candidate species to examine in more detail for confirmatory morphology, ecology or other sources of data that either support or refute the molecularly-defined species (Kekkonen and Hebert, 2014). In comparison, species that are inconsistently delimited, while also requiring an integrative taxonomy approach, are good candidates for more fine-scaled analyses, such as quantitative morphometrics (Lumley and Sperling, 2010; Schwarzfeld and Sperling, 2014) or population genetics (Shaffer and Thomson, 2007; Lumley and Sperling, 2011).

5. Conclusion

Based on morphology, Gauld (1985) estimated a fauna of approximately 50 Nearctic species of Ophion. This study indicates that the true number of Nearctic species is considerably higher and with more complete sampling will easily double that estimate. The 17 described species of Ophion therefore represent just a minute fraction of the total fauna. Similarly, even for the relatively well-studied British fauna, there appear to be a surprising number of previously unrecognized species. It is widely acknowledged that there are vast numbers of undescribed species among tropical, small-bodied taxa (e.g. Quickie, 2012). However, this study shows that even in a large-bodied, primarily temperate genus that is
readily attracted to lights and easily collected in large numbers, there is a wealth of diversity waiting to be discovered.

Acknowledgments

We are very grateful to Gavin Broad for loaning and donating British Ophion, for sending Ophion legs for sequencing, and for freely sharing his knowledge of British Ophion. Thanks also to Mark Shaw for sending a leg from O. forticoris. We wish to thank Jeremy deWaard for providing many specimens as well as nearly 100 COI sequences through BOLD. This project would not have been possible without the many collectors who provided specimens, especially: John Acorn, John Adams, Gary Anweiler, Charley Bird, Heather Bird, Brett Bodeux, Bryan Brunet, Jason Dombroskie, Julian Dupuis, Jessica Edwards, Dave Holden, Wes Hunting, Dave Langor, Nina Laney, Dave Lawrie, Lisa Lumley, Doug Macaulay, Boyd Mori, Ted Pike, Greg Pohl, and Chris Schmidt. We would also like to thank Martin Jones for assistance with jMOTU and Matthias Wolf for help regarding 4Sale and ProfDistS. Thank you to Seraina Klopfstein and two anonymous reviewers for their constructive comments on an earlier version of this manuscript. This work was supported by: the Sustainable Forest Management Network; Alberta Conservation Association Grants in Biodiversity; Natural Sciences and Engineering Research Council (NSERC) and Alberta Ingenuity (Alberta Innovates) student scholarships to MDS; and an NSERC Discovery grant to FAHS.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2015.08.003.

References

Gauld, I.D., 1985. The phylogeny, classification and evolution of parasitic wasps of