Molecular Phylogenetics and Evolution xxx (2011) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Species identification in the taxonomically neglected, highly diverse, neotropical parasitoid wasp genus *Notiospathius* (Braconidae: Doryctinae) based on an integrative molecular and morphological approach

Fadia Sara Ceccarelli [a], Michael J. Sharkey [b], Alejandro Zaldívar-Riverón [a,*]

[a] Colección Nacional de Insectos, Instituto de Biología, Universidad Nacional Autónoma de México, 3er. circuito exterior s/n, Cd. Universitaria, Copilco Coyoacán, A. P. 70-233, C. P. 04510. D. F., Mexico
[b] Department of Entomology, University of Kentucky, S-225 Ag. Sci. Building-N, Lexington, KY 40546-0091, USA

ARTICLE INFO

Article history:
Received 22 June 2011
Revised 24 October 2011
Accepted 24 October 2011
Available online xxxx

Keywords:
Species delineation
GMYC model
Doryctinae
Integrative taxonomy

ABSTRACT

Various DNA sequence-based methods for species delineation have recently been developed to assess the species-richness of highly diverse, neglected invertebrate taxa. These methods, however, need to be tested under a variety of conditions, including the use of different markers and parameters. Here, we explored the species diversity of a species-rich group of braconid parasitoid wasps, the Neotropical genus *Notiospathius*, including 233 specimens from 10 different countries. We examined sequences of two mitochondrial (mt) (COI, cyt *b*) and one nuclear (*wg*) gene fragments. We analysed them separately as well as concatenating the mt data with the general mixed Yule-coalescent (GMYC) model for species delineation using different tree-building methods and parameters for reconstructing ultrametric trees. We evaluated the performance of GMYC analyses by comparing their species delineations with our morphospecies identifications. Reconstructing ultrametric trees with a relaxed lognormal clock rate using the program BEAST gave the most congruent results with morphology for the two mt markers. A tree obtained with *wg* using the programs MrBayes + Pathd8 had the fewest cases of incongruence with morphology, though the performance of this nuclear marker was considerably lower than that of COI and cyt *b*. Species delimitation using the coalescent prior to obtain ultrametric trees was morphologically more congruent with COI, whereas the Yule prior was more congruent with cyt *b*. The analyses concatenating the mt datasets failed to recover some species supported both by morphology and the separate analyses of the mt markers. The highest morphological congruence was obtained with the GMYC analysis on an ultrametric tree reconstructed with cyt *b* using the relaxed lognormal clock rate and the Yule prior, thus supporting the importance of using alternative markers when the information of the barcoding locus (COI) is not concordant with morphological evidence. Seventy-one species were delimited based on the congruence found among COI, cyt *b* and morphology. Both mt markers also revealed the existence of seven potential cryptic species. This high species richness from a scattered geographical sampling indicates that there is a remarkable number of *Notiospathius* species that remains undiscovered.

1. Introduction

The species as a taxonomic unit has long been used for measuring the biodiversity and abundance of living organisms in ecosystems (Gaston, 2000). Many groups of organisms, however, only have a small fraction of their species described. This is especially the case in highly diverse invertebrate groups, for which it would be virtually impossible to know the number of species in the near future employing morphology-based studies alone. The recent introduction of standardised DNA sequence fragments (Hebert et al., 2003a; Lahaye et al., 2008; Monaghan et al., 2005) together with the development of DNA sequence-based methods for species delineation (Blaxter, 2004; Pons et al., 2006; Monaghan et al., 2009; Yang and Rannala, 2010) are proving to be robust tools to help unveil biodiversity in a fast, accurate way. In particular, a ~650 base-pair fragment of the mitochondrial (mt) cytochrome oxidase I (COI) gene has rapidly been adopted by the scientific community as the "barcoding" locus for sequence-based identification of many animal groups (Hebert et al., 2003b; Hajibabaei et al., 2007; Miller, 2007; Janzen et al., 2009).

Barcoding analyses for species delineation are frequently carried out using a Neighbour-Joining (NJ) distance analysis that graphically displays divergences among sequences. This phenogram is composed of clusters of sequences, which are considered

* Corresponding author. Fax: +52 55 55 50 01 64.
E-mail address: azaldivar@ibiologia.unam.mx (A. Zaldívar-Riverón).

1055-7903/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2011.10.018

as "barcoding" species based on whether they are separated by previously defined divergence distances (e.g. Hebert et al., 2004; Smith et al., 2006; Ward et al., 2005). This has proved to be a valuable procedure in cases in which there are too many species to be identified within a limited timeframe (e.g. Smith et al., 2009), for detecting cryptic species (e.g. Burns et al., 2008a; Hebert et al., 2004; Pauls et al., 2010; Smith et al., 2008), for associating morphologically different semaphoronts (e.g. Miller et al., 2005; Gattolliat and Monaghan, 2010; Pinzon-Navarro et al., 2010) or for associating sexes within a species (e.g. Sheffield et al., 2009). These analyses, however, have been criticised since they employ arbitrary sequence divergence criteria for delimiting species. Establishing sequence divergence values is based on the premise that nucleotide substitution rates follow a fixed clock-like pattern. This is problematic since it is now generally accepted that rates can vary over time and across taxa (Drummond et al., 2006). In insects, for example, molecular clocks have often been used arbitrarily, without taking varying substitution rates between markers and taxa into account (Papadopoulou et al., 2010).

The recent development of rigorous, quantitative sequence-based methods for species delineation is changing the way biodiversity is studied, since they take into account theory concerning processes that have led to speciation, e.g. reproductive isolation, independent limitation, and divergent selection (Barraclough, 2010). These novel methods, however, need to be tested under a variety of situations, and the effects of different markers and parameters on accuracy must be scrutinised. Various approaches have been created to test for independent limitation (e.g. Meyer and Paulay, 2005; Blaxter et al., 2005; Hong et al., 2006; Pons et al., 2006), that is, the separate coalescence of different populations, which results in distinct clusters of sequences separated from each other by longer internal branches (Barraclough et al., 2003; Birky et al., 2005). One of these approaches, the general mixed Yule-coalescent (GMYC) method (Pons et al., 2006; Fontaneto et al., 2007), seeks to identify independently evolved lineages (species) as the most likely point of transition from coalescent to speciation branching patterns on an ultrametric phylogenetic tree with branch lengths scaled to time. In a recent study, this method gave accurate species boundaries estimates using three different mt markers in several highly diverse, taxonomically neglected insect taxa from Madagascar (Monaghan et al., 2009).

Since the GMYC method requires an ultrametric tree with branching rates based on clock-like assumptions, selecting the appropriate priors and branching pattern parameters for tree reconstruction is critical for delimiting species accurately. Various programs have been developed to estimate "relaxed" molecular clocks (e.g. r8s: Sanderson, 2003; multidivide: Thorne and Kishino, 2002; Pathd8: Britton et al., 2007; BEAST: Drummond and Rambaut, 2007), most of which require a previously built phylogram from which the divergence time estimates are calculated. Of these programs, BEAST is the only one based on a Bayesian method that simultaneously reconstructs a phylogeny with branch lengths that can be scaled to time. Two uncorrelated, relaxed clock branch rate models are available in this program, the lognormal and the exponential models; the first one assumes that rate changes occur along branches, and the second assumes that rates change at the nodes of the tree independent of their branch lengths (Drummond et al., 2006; Drummond and Rambaut, 2007). In addition, the random local clock model applies different rates to sub-sections of the trees (Drummond and Suchard, 2010). BEAST can also implement different tree priors, with the coalescent and Yule priors being used for population-level and species-level data, respectively (Drummond et al., 2006).

The Neotropical braconid parasitoid wasp genus *Notiospathius* Matthews & Marsh (Doryctinae) represents a prime example of a species-rich, morphologically conserved insect taxon. Only 32 species have been described for this genus to date (Zaldívar-Riverón and De Jesus-Bonilla, 2010; De Jesús-Bonilla et al., 2011), although its actual diversity is much higher, with a large, undetermined number of species distributed from northern Mexico to northern Argentina (Zaldívar-Riverón, unpublished data). Species of *Notiospathius* appear to be concentrated in tropical, lowland regions, though some species have also been recorded in more elevated, temperate zones (Zaldívar-Riverón et al., unpublished data).

In this study we investigate the species boundaries of *Notiospathius* collected in several different localities. We applied GMYC analyses employing two mt [COI and cytochrome *b* (cyt *b*)] genes and one nuclear (*wg*) marker separately; we also concatenated the two mt gene fragments, and tested a variety of methods and parameters for reconstructing ultrametric trees with branch lengths scaled to time. We then evaluated the performance of the datasets examined and of the sets of methods and parameters applied for tree reconstruction in the GMYC analyses by comparing their species boundary estimates with our morphospecies identifications.

2. Materials and methods

2.1. Specimens sampling

A total of 270 specimens of *Notiospathius* were obtained from collecting excursions carried out from 1993 to 2010 in several localities spread throughout 10 Neotropical countries, with many of the sampled localities only represented by one or two specimens. Of these we were able to generate sequences for 216 specimens (see Supplementary material – S1). We follow De Jesús-Bonilla et al.'s (2011) taxonomic change and consider *Hansonorum* as a junior synonym of *Notiospathius*. Species previously assigned to *Hansonorum* were distinguished from those of *Notiospathius* by having a basal tubercle on the hind coxa; the presence/absence of this structure, however, was shown to be highly homoplastic within Doryctinae (Zaldívar-Riverón et al., 2007, 2008). We included four specimens belonging to the small genera *Masonius* (two specimens) and *Tarasco* (two specimens), which are currently represented by one and three described species, respectively. These two genera appeared intermingled in a clade with species of *Notiospathius* in previous molecular phylogenetic studies (Zaldívar-Riverón et al., 2007, 2008), suggesting paraphyly of the latter genus. COI sequences of twelve additional specimens (10 *Notiospathius* and two *Masonius*) from a previous study (Zaldívar-Riverón et al., 2010) were also included. In addition, one specimen of *Spathius* was used for rooting all phylogenetic trees. Specimens were preserved in 95% to absolute ethanol until they were subsequently processed for DNA sequencing. A list of specimens included in this study, their voucher numbers, locality details and GenBank accession numbers is in the Supplementary material (S1). Specimens are deposited in the Colección Nacional de Insectos, Instituto de Biología, Universidad Nacional Autónoma de México and the Hymenoptera Institute Collection, Lexington, Kentucky.

2.2. DNA sequencing

We generated sequences of two mt markers: 529 base-pairs (bp) of the COI gene, which comprises most of the barcoding region, and a 344 bp fragment belonging to the cyt *b* gene. These are two of the most widely used mt genes for phylogenetic analyses of closely related taxa, as well as for phylogeographic and population genetic studies in animals, especially insects (Lin and Danforth, 2004; Simmons and Weller, 2001; Simon et al., 1994). A fragment (410 bp) of the nuclear wingless (*wg*) gene was also sequenced. DNA was extracted for all the collected specimens with

Please cite this article in press as: Ceccarelli, F.S., et al. Species identification in the taxonomically neglected, highly diverse, neotropical parasitoid wasp genus *Notiospathius* (Braconidae: Doryctinae) based on an integrative molecular and morphological approach. Mol. Phylogenet. Evol. (2011), doi:10.1016/j.ympev.2011.10.018

the DNeasy blood and tissue extraction kit (QIAGEN®) using a non-destructive technique, where a whole individual is left to digest overnight in 100 μl of QIAGEN's ATL buffer and 20 μl of proteinase K. The liquid is then passed through the kit's columns following the manufacturer's instructions. Specimens were washed twice with distilled water after being incubated and were subsequently dried, mounted, and labelled.

Amplifications of the selected DNA fragments were prepared in a total volume of 25 μl, with 2.5 μl of 10 × PCR buffer, 1 μl of $MgCl_2$, 0.25 mM of each dNTP, 0.4 μM of each primer, 0.2 μl of platinum Taq polymerase (Invitrogen®), 5 μl of DNA template and 12 μl of ddH_2O. PCRs were carried out using the primers co1NotF1 (5′-TAG AAT TAG GTA TAC CDG G-3′) (newly developed) and LEP-R1 (5′-TAA ACT TCT GGA TGT CCA AAA A-3′) (Hebert et al., 2004) for the COI fragment, Cyt *b* fwd (5′-TCT TTT TGA GGA GCW ACW GTW ATT AC-3′) and Cyt *b* rev (5′-AAT TGA ACG TAA AAT WGT RTA AGC AA-3′) (Belshaw and Quicke, 1997) for the cyt *b* fragment, and Wg_1 (5′-GAR TGY AAR TGY CAY GGY ATG TCT GG-3′) and Wg_2 (5′-ACT ICG CRC ACC ART GGA ATG TRC A-3′) (Brower and DeSalle, 1998) for the *wg* gene fragment. PCR cycling conditions were as follows: for COI, an initial denaturing step of 1 min at 94 °C, followed by 35 cycles of 94 °C for 30 s, 48 °C for 45 s and 72 °C for 1 min, and a final extension step at 72 °C for 7 min; for cyt *b* and *wg* an initial denaturing step of 2 min at 95 °C followed by 35 cycles of 95 °C for 1 min, 54 °C for 1 min and 72 °C for 1 min 30 s, and a final extension step at 72 °C for 10 min. PCR products were either purified using Millipore columns (Millipore®) and sequenced on an ABI Prism 3100 Genetic Analyzer, or sent directly to the High-Throughput Genomics Unit at the University of Washington (http://www.htseq.org/index.html).

All sequences were edited using Sequencher 4.1.4 (Gene Codes Corp.) and aligned by eye. A few of the COI and cyt *b* sequences were slightly length variable and thus their alignments were confirmed by translating their nucleotides to amino acids with the program MacClade version 4.06 (Maddison and Maddison, 2000). The *wg* sequences were not length-variable and therefore their alignment was straightforward.

2.3. Phylogenetic reconstructions and species boundaries analyses

We first conducted separate sequence-based analyses employing the GMYC method for the mt and nuclear markers. For several of the older specimens, sequences could not be obtained due to DNA degradation, and therefore there was a considerable mismatch between the COI, cyt *b* and *wg* datasets generated. For comparative purposes, we therefore created two datasets for each mt marker: one for which sequences of a given specimen were obtained for both mt markers (COI_{common}, cyt b_{common}), and the other with all available sequences for each marker (COI_{all} and cyt b_{all}). In addition, two more datasets were created, one concatenating the mt sequence data of specimens for which both mt gene fragments were obtained ($both_{common}$), and another containing all mt sequences generated for both mt markers ($both_{all}$), considering those sequences that could not be obtained for one of the markers as missing data. The COI_{common} and cyt b_{common} datasets were first run with the sets of methods and parameters outlined below in order to generate ultrametric trees for the GMYC analyses that yielded the most congruent species delineation based on comparisons with morphospecies identification. In contrast, the two concatenated datasets ($both_{common}$, $both_{all}$) were only run with the four sets of parameters and methods that yielded species delineations that were most congruent with morphology in the separate GMYC analyses performed with COI and cyt *b*. The number of *wg* sequences obtained was considerably lower than those of COI and cyt *b*. We therefore only ran the sets of methods and parameters for ultrametric tree construction (detailed below) with the complete *wg* dataset.

One set of methods first executed a partitioned MCMC Bayesian analysis with MrBayes version 3.1.2 (Ronquist and Huelsenbeck, 2003). This analysis consisted of two independent runs of 20,000,000 generations each, sampling trees every 1000 generations. All datasets were partitioned into first, second and third codon positions and the appropriate nucleotide substitution model [GTR + Γ + I (Lanave et al., 1984) for the three codon positions of the mt gene fragments; GTR + Γ for the 1st and 3rd, and Kimura 2-parameter (Kimura, 1980) for the 2nd codon position of the *wg* gene fragment], chosen with MrModeltest 2.3 (Nylander, 2004) in PAUP version 4.0b (Swofford, 2001), was applied to each partition. A majority consensus tree with branch lengths was reconstructed for each run after discarding the first 15,000 sampled trees. The Bayesian phylograms were then passed through the program Pathd8 (Britton et al., 2007) to obtain an ultrametric tree with branch lengths using the MPL (Mean Path Length) method (Britton et al., 2002). The same phylograms were also passed through the program r8s version 1.71 (Sanderson, 2003), using the penalised likelihood (PL) method (Sanderson, 2002) and the smoothing factor was established by first running a cross-validation test on the input files.

In addition, the separate mt and nuclear datasets were analysed using Bayesian methods with the program BEAST version 1.6.1 (Drummond and Rambaut, 2007), running MCMC chains for 20 million generations, sampling trees every 1000 generations, and using the GTR + Γ + I substitution model in all matrices but without partitions. We conducted six relaxed clock analyses by combining three branching rate clock models [lognormal (−l), exponential (−e) and random local (−rlc)] clock models, with two tree priors [constant Yule's speciation (−y) and coalescent (−c)] in order to identify which combination yielded the results that were most congruent with morphology. The last 1000 sampled trees obtained in each analysis were employed to reconstruct a maximum clade credibility tree with the program TreeAnnotator version1.6.1 (part of the BEAST package, Drummond and Rambaut, 2007), using the mean heights option.

The ultrametric phylogenies recovered with the MrBayes/Pathd8, MrBayes/r8s and BEAST analyses were subjected to GMYC analyses implemented in the SPLITS package (available from http://r-forge.r-project.org/projects/splits/) with the program R 2.10.1 (R Core Development Team, 2009). A list of delimited "GMYC species" (described in the program's output as ML entities) was compiled from the graphical output of the GMYC analysis in R.

2.4. Comparisons between GMYC species and morphospecies

Specimens for which we were able to generate a sequence of at least one marker were morphologically examined in order to compare the number of morphospecies distinguished with the number of GMYC species recovered in each of the above analyses. Morphospecies discrimination was performed based on examination of 31 diagnostic characters that were selected from the taxonomic literature (Matthews and Marsh, 1973; Marsh, 2002; De Jesús-Bonilla et al., 2011). A table with our discriminated morphospecies and the 31 characters scored for all the individuals that were sequenced in this study can be found in the Supplementary material (S2).

For the separate and combined data sets, we selected the results that were most congruent with our morphospecies identifications. We measured this congruence as the lowest number of morphospecies that were fused, divided, or whose specimens were placed in alternative groupings (termed "misplaced") with respect to the GMYC species. A misplaced morphospecies was identified as one having one or more specimens placed in a clade together with

Please cite this article in press as: Ceccarelli, F.S., et al. Species identification in the taxonomically neglected, highly diverse, neotropical parasitoid wasp genus *Notiospathius* (Braconidae: Doryctinae) based on an integrative molecular and morphological approach. Mol. Phylogenet. Evol. (2011), doi:10.1016/j.ympev.2011.10.018

specimens belonging to a different morphospecies. We considered misplaced and fused morphospecies in the GMYC analyses as a failure of the markers to identify morphologically distinguishable species. The presence of divided morphospecies into two or more GMYC species on the other hand was considered to represent potential cryptic or pseudocryptic (*sensu* Sáez and Lozano, 2005) species. We followed a conservative approach and preferred the GMYC analyses that yielded the lowest number of cryptic species. The analyses with the sets of methods and parameters for the mt datasets that were most congruent with morphology were also employed in GMYC analyses with the COI_{all}, cyt b_{all}, $both_{common}$ and $both_{all}$ datasets; this allowed us to explore the species diversity contained in all of our sequenced specimens.

3. Results

3.1. Sequence divergence

We obtained 171 COI, 160 cyt *b* and 121 *wg* sequences from a total of 233 individuals assigned to *Notiospathius* (225 specimens), *Masonius* (four specimens), *Tarasco* (two specimens) and *Spathius* (one specimen, outgroup). The rate of sequencing success generally depended of the age of preservation of the specimens. The maximum age of specimens preserved in ethanol from which we were able to obtain sequences was 17 years. The DNA sequence variation of the mt and nuclear datasets examined is summarised in Table 1. All sequences were considered for the different analyses performed, regardless of the fact that several of them had identical haplotypes. The COI fragment had more base-pairs than either the cyt *b* or the *wg* gene fragments, and it also contained the highest number of variable sites and unique haplotypes (COI_{common} = 106; cyt b_{common} = 100; *wg* = 81). However, on a per-base-pair level, the COI_{common}, cyt b_{common} and *wg* sites are variable respectively at 46.3%, 61.3% and 34.4%, indicating that *cyt b* has the highest nucleotidic substitution rate.

3.2. Species delimitation analyses

The results of the GMYC analyses performed for the *wg*, COI_{common}, cyt b_{common} and $both_{common}$ datasets are presented in Table 2. The threshold (T) values indicating the time of transition between coalescence and speciation branching patterns occurred on the ultrametric trees varied from −0.9 (MB + d8 *wg*) to −10.6 (MrBayes + r8s COI_{common}). These considerable differences of T values are not comparable due to the distinct scale values given to the relative ages on the ultrametric trees generated by the programs BEAST, r8s and Pathd8.

All the analyses yielded a high number of GMYC species represented by a single individual. The number of GMYC species recovered with the different methods and parameters employed for reconstructing ultrametric trees ranged from 60 (MrBayes + Pathd8) to 75 (BEAST-rlc-y) for COI_{common}, from 69 (BEAST-e-y) to 87 (BEAST-rlc-y) for cyt b_{common}, from 46 (MrBayes + r8s) to 113 (BEAST-rlc-y) for *wg*, and from 77 (BEAST-l-c) to 81 (MrBayes + r8s) for $both_{common}$ (Table 2). The range of confidence interval differences for the number of GMYC species delineated with the eight different sets of parameters and methods was generally narrower for COI_{common} (11–35), followed by cyt b_{common} (10–74) and *wg* (10–119). The confidence interval differences for the number of GMYC species delineated with the four different sets of parameters and methods applied to the $both_{common}$ dataset ranged from nine to 21.

Table 1
Sequence variation for the five datasets examined.

Data sets	*n*	Base pairs	H	VS	PIS
COI_{common}	123	529	107	245	216
cyt b_{common}	123	344	101	211	182
COI_{all}	183	529	145	253	221
cyt b_{all}	160	344	129	217	188
Wg	121	410	81	141	118
$both_{common}$	123	873	–	456	398
$both_{all}$	220	873	–	470	409

n = number of sequences; H = number of haplotypes; VS = variable sites; PIS = parsimoniously informative sites.

Table 2
Number of GMYC "species" recovered and outputs obtained from the single-threshold GMYC analyses performed for the two mt markers examined.

	T	NC	NS	MS	CI	L_0	L_{GMYC}	LR
COI_{common}				65				
e-c	−0.05	20	68		66–82	555.1	564.8	19.3*
e-y	−0.46	22	71		68–103	313.2	317.4	8.4
l-c	−0.13	19	69		61–78	574.2	603.3	58.2*
l-y	−0.2	19	69		64–82	358.2	390.5	64.6*
rlc-c	−0.02	22	65		61–78	652.7	678.9	52.4*
rlc-y	−0.16	20	75		62–82	367.2	392.8	51.2*
MB + d8	−0.07	18	60		59–75	538.8	563.7	49.9*
MB + r8s	−10.6	19	67		62–73	124.3	91.7	65.2*
cyt b_{common}				65				
e-c	−0.05	20	79		70–104	505.0	509.7	9.6
e-y	−0.18	19	86		80–100	374.3	378.9	9.2
l-c	−0.3	20	74		61–87	295.4	306.0	21.3*
l-y	−0.2	23	73		62–86	341.1	362.1	41.9*
rlc-c	−0.02	24	79		75–85	579.1	594.7	31.0*
rlc-y	−0.24	23	69		61–81	383.9	398.9	30.0*
MB + d8	−0.05	23	71		2–76	530.4	564.2	67.5*
MB + r8s	−0.1	22	69		60–77	468.5	488.5	40.0*
wg				54				
e-c	−0.01	9	109		2–120	501.0	503.0	4.0
e-y	−0.11	10	108		1–115	336.7	338.0	2.5
l-c	−0.01	20	71		67–77	735.9	742.2	12.6*
l-y	−0.11	11	105		1–115	331.5	332.9	2.7
rlc-c	−0.001	9	102		1–120	811.6	813.1	3.0
rlc-y	−0.004	6	113		1–120	479.0	480.0	2.4
MB + d8	−0.9	20	51		33–62	550.9	554.2	6.7
MB + r8s	−0.42	24	46		42–65	262.5	266.2	7.5
$both_{common}$				65				
l-c	−0.06	22	77		67–88	483.5	507.9	48.8*
l-y	−0.15	21	79		73–87	366.2	394.1	56.0*
MB + d8	−0.03	23	79		69–84	526.7	567.1	80.9*
MB + r8s	−0.03	22	81		76–85	541.1	568.1	54.1*

Ultrametric tree reconstructions with the following parameters: e-c/e-y = relaxed exponential clock rates using coalescent or Yule tree priors in BEAST, respectively; l-c/l-y = relaxed lognormal clock rates using coalescent or Yule tree priors in BEAST, respectively; rlc-c/rlc-y = random local clock rates using coalescent or Yule tree priors in BEAST, respectively; MB + d8 = MrBayes + Pathd8; MB + r8s = MrBayes + r8s; T = threshold genetic distance from the branch tips where the coalescent-speciation transition occurred; NC = number of clusters (GMYC "species" with more than one individual); NS = number of GMYC "species" discriminated; MS = number of morphospecies identified from 31 selected external morphological characters; CI = confidence intervals of GMYC "species"; L_0 = likelihood of null model; L_{GMYC} = likelihood of GMYC model; LR = likelihood ratio with significance indicated by an asterisk.
* $p < 0.01$.

3.3. Molecular versus morphological species delimitations

The 233 individuals examined in this study were placed in a total of 100 morphospecies (see Supplementary material S2). The 121 *wg*, 171 COI_{all} and 160 cyt b_{all} sequences belonged to 54, 86 and 76 morphospecies, respectively. Sixty-five of the 100 discriminated morphospecies have sequences both for COI and cyt *b*, whereas 95 of them have sequences for at least one of the two mt markers. A comparison of the levels of incongruence between our delimited morphospecies and the GMYC species is given in Table 3.

Table 3
Levels of incongruence observed between species delimited with the different GMYC analyses and our morphospecies identifications based on examination of 31 selected external morphological characters.

Marker/matrix	Incongruence						Quantitative difference
	Method	Split	Fused	Misplaced	Total	% Differences	% **Increase**/*decrease*
COI_{common}	e-c	6	1	8	15	21.7	**4.6**
	e-y	9	1	7	17	23.6	**9.2**
	l-c	6	1	6	13	18.8	**6.2**
	l-y	6	0	9	15	21.4	**6.2**
	rlc-c	5	3	8	16	24.6	0
	rlc-y	11	0	7	18	24.0	**15.4**
	MB + d8	2	5	8	15	25.0	*7.7*
	MB + r8s	6	1	9	16	23.9	**3.1**
	Average (+/− s.d.)	6.4 (2.7)	1.5 (1.7)	7.8 (1.0)	15.6 (1.5)	22.9 (2.1)	**4.6** (6.7)
cyt b_{common}	e-c	14	0	0	14	17.5	**21.5**
	e-y	21	0	0	21	24.1	**32.3**
	l-c	10	1	0	11	14.7	**13.8**
	l-y	6	0	0	6	8.2	**12.3**
	rlc-c	13	0	0	13	16.5	**21.5**
	rlc-y	6	3	0	9	13.0	**6.1**
	MB + d8	6	1	1	8	11.3	**9.2**
	MB + r8s	6	3	1	10	14.5	**6.2**
	Average (+/− s.d.)	10.3 (5.5)	1 (1.3)	0.3 (0.5)	11.5 (4.6)	15.0 (4.7)	**15.4** (9.1)
wg	e-c	55	0	0	55	50.5	**101.9**
	e-y	54	0	1	55	50.9	**100**
	l-c	18	0	2	20	28.7	**31.5**
	l-y	51	0	0	51	48.6	**94.4**
	rlc-c	48	0	0	48	47.1	**88.9**
	rlc-y	59	0	0	59	52.2	**109.3**
	MB + d8	5	8	0	13	25.5	*5.6*
	MB + r8s	6	14	1	21	45.7	*14.8*
	Average (+/− s.d.)	37 (23.2)	2.8 (5.3)	0.5 (0.8)	40.3 (18.8)	43.7 (10.5)	**63.2** (51.3)
$both_{common}$	l-c	12	0	0	12	15.6	**18.5**
	l-y	13	0	0	13	16.5	**21.5**
	MB + d8	13	0	0	13	16.5	**21.5**
	MB + r8s	15	0	0	15	18.5	**24.6**
	Average (+/− s.d.)	13.3 (1.3)	0 (0)	0 (0)	13.3 (1.3)	16.8 (1.2)	**21.5** (2.5)

Ultrametric tree reconstructions with the following parameters: e-c/e-y = relaxed exponential clock rates using coalescent or Yule tree priors in BEAST, respectively; l-c/l-y = relaxed lognormal clock rates using coalescent or Yule tree priors in BEAST, respectively; rlc-c/rlc-*y* = random local clock rates using coalescent or Yule tree priors in BEAST, respectively; MB + d8 = MrBayes + Pathd8; MB + r8r = MrBayes + r8s; Average (+/− standard deviation) = average of differences for all methods for each marker; split = number of morphospecies containing two or more GMYC species; fused = number of morphospecies merging two or more GMYC "species"; misplaced = morphospecies whose specimens appeared placed in separate sequence clusters; total = total number of incongruent species between morphology and GMYC analyses; % differences = percentage of incongruent species between morphology and GMYC analyses ((total differences/GMYC "species") × 100); quantitative difference = percentage of increase/decrease of GMYC species with regards to number of morphospecies.

All but four of the analyses recovered more GMYC species than the morphospecies included in each dataset (54 morphospecies for *wg*, and 65 for the separate mt and $both_{common}$ datasets, respectively). Overall, the levels of incongruence between morphospecies and GMYC species were lower for cyt b_{common} than for the remaining separate and concatenated datasets. The set of methods and parameters with the lowest incongruence with morphology for each marker were BEAST-l-c for COI_{common} (incongruence = 18.8%), BEAST-l-y for cyt b_{common} (incongruence = 8.2%) and MrBayes + Pathd8 for *wg* (incongruence = 25.5%). Incongruence in the BEAST-l-c COI_{common} analysis was composed of six split, one fused and seven misplaced morphospecies, in the BEAST-l-y cyt b_{common} analysis of six fused morphospecies and in the MrBayes + Pathd8 *wg* analysis of five split and eight fused morphospecies. These analyses recovered 69 [COI_{common}: confidence interval (CI) = 61–78, LR = 58.2; $p < 0.001$], 73 (cyt b_{common}: CI = 62–86, LR = 41.9; $p < 0.001$) and 51 (*wg*: CI = 33–62, LR = 6.69; $p = 0.08$) GMYC species, respectively. Of the four sets of methods and parameters applied to the COI + cyt *b* dataset ($both_{common}$), the BEAST-l-c analysis was the most congruent with morphology (incongruence = 18.6%; CI = 67–88, LR = 48.8; $p < 0.01$), yielding an ultrametric tree that resulted in 77 delineated GMYC species. The number of GMYC species constituted by sequence clusters in the above separate and combined analyses ranged from 19 (BEAST-l-c COI_{common}) to 23 (BEAST-l-y cyt b_{common}).

Fig. 1 shows a Bayesian ultrametric tree ($both_{all}$; MrBayes + Pathd8) that summarises the congruence/incongruence found between the morphospecies identified and the GMYC species recovered by the COI_{all}, cyt b_{all} and *wg* datasets using the selected sets of parameters and methods. A list with the GMYC species recovered for each marker/method/parameter combination can be found in the Supplementary material (S3). The ultrametric trees that yielded the lowest levels of incongruence for COI_{common} (BEAST-l-c), cyt b_{common} (BEAST-l-y), $both_{common}$ (BEAST-l-c) and *wg* (MrBayes + Pathd8) and their GMYC species obtained are shown in the Supplementary materials S4–S7, respectively. The trees recovered with the remaining analyses performed can be provided by the authors upon request.

When COI_{all} and cyt b_{all} were analysed with the sets of methods and parameters that gave the most congruent results with morphology they recovered 94 and 88 GMYC species (COI_{all}: CI = 86–99, LR = 104.0; cyt b_{all}: CI = 85–106, LR = 46.7; $p < 0.001$ in both cases), respectively, 64 and 67% of which were singletons. Levels of incongruence with morphology in the latter analyses were of 24.4% for COI_{all} and of 17.1% for cyt b_{all}, and their incongruence consisted of 12 split and nine misplaced and of 12 split and one fused morphospecies, respectively. The concatenated mt dataset with missing data ($both_{all}$) had 220 terminal taxa, of which 97 (44.1%) only had sequences for one or the other mt gene. The ultrametric tree constructed using BEAST-l-c for $both_{all}$ obtained 107

Please cite this article in press as: Ceccarelli, F.S., et al. Species identification in the taxonomically neglected, highly diverse, neotropical parasitoid wasp genus *Notiospathius* (Braconidae: Doryctinae) based on an integrative molecular and morphological approach. Mol. Phylogenet. Evol. (2011), doi:10.1016/j.ympev.2011.10.018

GMYC species (CI = 97–118, LR = 88.2; $p < 0.01$), of which 60.7% were singletons. The incongruence between this concatenated dataset and morphology was higher than that obtained in the separate analyses of the mt datasets (25.2%; 14 split, four fused and nine misplaced morphospecies). Moreover, the GMYC analysis of $both_{all}$ maintained most of the misplaced morphospecies also found in the separate analyses of the COI marker. The ultrametric trees reconstructed for COI_{all}, cyt b_{all} and $both_{all}$ and their GMYC species recovered can be found in the Supplementary materials S8–S10, respectively.

Only 15 species were supported by COI_{all}, cyt b_{all}, *wg* and morphology, this due to the high incongruence of species recovered with *wg* compared to those obtained with the mt markers and morphology and by the lower number of sequences generated for this nuclear marker. A total of 64 and 55 species were supported by COI_{all} and morphology and by Cyt b_{all} and morphology, respectively. Of these species, 16 contained specimens that were only sequenced for COI and seven only for cyt *b*, whereas 48 were supported by morphology and the two mt markers. The number of concordant species in the concatenated mt datasets and morphology was 54 and 66 for $both_{common}$ and $both_{all}$, respectively. Some species supported by COI_{all} and morphology and cyt b_{all} and morphology, however, were not supported by $both_{all}$ and morphology (see Supplementary material S3). A list with the species supported by morphology and by all the GMYC analyses performed is given in Supplementary material (S3).

Seven additional GMYC species were recovered by our selected COI_{common} and cyt b_{common} analyses but not by morphology (see Supplementary material S3). Two of these GMYC species were merged in a single morphospecies having individuals from different, isolated localities in Colombia. One of these GMYC species has specimens from the southwest (department of Valle de Cauca, voucher Nos. CNIN 430, 530) and the other from the southeast (department of Vaupés, voucher Nos. 536) of the country. A third GMYC species contains specimens from northern Venezuela (voucher Nos. CNIN 362, 367, 382, 561, 564), which were grouped in one morphospecies together with individuals from the same region and Trinidad and Tobago whose GMYC species were not supported by both mt markers. Finally, the remaining four GMYC species have specimens from two localities in Mexico corresponding to two morphospecies, two having individuals from Los Tuxtlas, Veracruz (voucher Nos. CNIN 389 and CNIN 364, 634 and 635, respectively), and two with specimens from Sierra de Juárez, Oaxaca (voucher Nos. CNIN 448, 452, 457–58, 462–64, 468–71, 479 and CNIN 379–80, 446–47, 453, 459, 461). We examined additional external morphological features others than the 31 originally selected for the specimens of the above seven GMYC species supported by both mt markers but did not find consistent variation differentiating them.

4. Discussion

4.1. Comparison of tree building methods and parameters in species delineation

Estimating divergence times accurately is a crucial step for delineating species employing the GMYC method. Currently, BEAST is becoming the most widely used program for calculating divergence time estimates since it simultaneously reconstructs phylogenetic relationships; it is also being the only program of its kind that allows the implementation of different branch rate models (Drummond et al., 2006). Our results agree with the use of BEAST for reconstructing the ultrametric trees scaled to time needed for delineating GMYC species because the results obtained were more congruent with morphology than those of the other programs examined (MrBayes, r8s, Pathd8).

We also found that the accuracy of the GMYC method is sensitive to the branch rate model implemented, since our trees obtained with the lognormal clock rate recovered more GMYC species that were congruent with morphology than the trees built with the other two branching rate models explored. Previous studies carried out with both simulated and real DNA sequence data have revealed that the lognormal rate model generally recovers more accurate clock estimates and phylogenetic relationships than the exponential rate model (Drummond et al., 2006; Monaghan et al., 2009; Wertheim et al., 2010). These findings are concordant with our GMYC analyses, which favour the ultrametric trees obtained using the lognormal clock rate, whereas the exponential clock rate often leads to species overestimation. Regarding the recently developed random local clock (Drummond and Suchard, 2010), which allows for different rates to be applied to sub-trees, it may not have brought any advantage to our datasets because the evolutionary rates among their lineages do not vary considerably.

The performance of the Yule and coalescence priors for estimating branch lengths in our analyses varied with respect to the marker examined, with the former one having the highest congruence with morphology using cyt *b* and the latter using COI and *wg*. The coalescent prior reconstructs large phylogenetic trees without altering the number of substitutions per site of its branch lengths by reducing the substitution rate accordingly, reason why this prior performs better with lower rates (Drummond et al., 2002). This explains the better performance of the coalescent prior in our study in analyses using the COI gene, which has a lower evolutionary rate than cyt *b*, and remarks the need to test both priors in order to obtain more accurate species delineations in GMYC analyses.

All our GMYC analyses recovered a high number of singletons, that is, species represented by a single sequence. Monaghan et al. (2009) mentioned that their species delineation analyses appeared to perform well despite the fact that approximately 60% of their GMYC species recovered were singletons. Similarly, most of our GMYC analyses were fairly congruent with our morphological identifications despite the high number of singletons that they recovered (e.g. 64% and 67% of singletons in the COI_{all} and cyt b_{all} analyses, respectively). Our results therefore confirm that a high proportion of singletons in GMYC analyses do not seem to negatively influence the performance of the model. In addition, it appears that the GMYC method performs well despite the inclusion of a number of haplotypes represented by two or more sequences.

4.2. Performance of mt and nuclear markers

Our species delineation analyses with cyt *b* were generally more congruent with our discriminated morphospecies than the analyses with COI. The cyt *b* gene has a higher nucleotide substitution rate than COI at species or lower levels in insects (Simmons and Weller, 2001). The former gene can therefore show variation where COI lacks it in recently derived species. This is supported by studies that revealed that the barcoding region sometimes has minimal, or even lacks, nucleotidic variation in morphologically distinct species (Burns et al., 2008b; McFadden et al., 2011), thus demonstrating the necessity of using faster evolving markers for resolving species limits in diverse, closely related taxa. Moreover, some of the species clusters recovered by COI were evidently misleading in light of our morphological identifications, which were instead congruent with the species delimited by cyt *b*. This difference in signal of the two mt markers, despite being from two linked genes, thus supports the importance of not relying on only one source of evidence in order to avoid incorrect phylogenetic inferences.

The nuclear marker employed in this work had the highest levels of incongruence with respect to morphological and mt data. Recently diverged species are expected to exhibit more

Please cite this article in press as: Ceccarelli, F.S., et al. Species identification in the taxonomically neglected, highly diverse, neotropical parasitoid wasp genus *Notiospathius* (Braconidae: Doryctinae) based on an integrative molecular and morphological approach. Mol. Phylogenet. Evol. (2011), doi:10.1016/j.ympev.2011.10.018

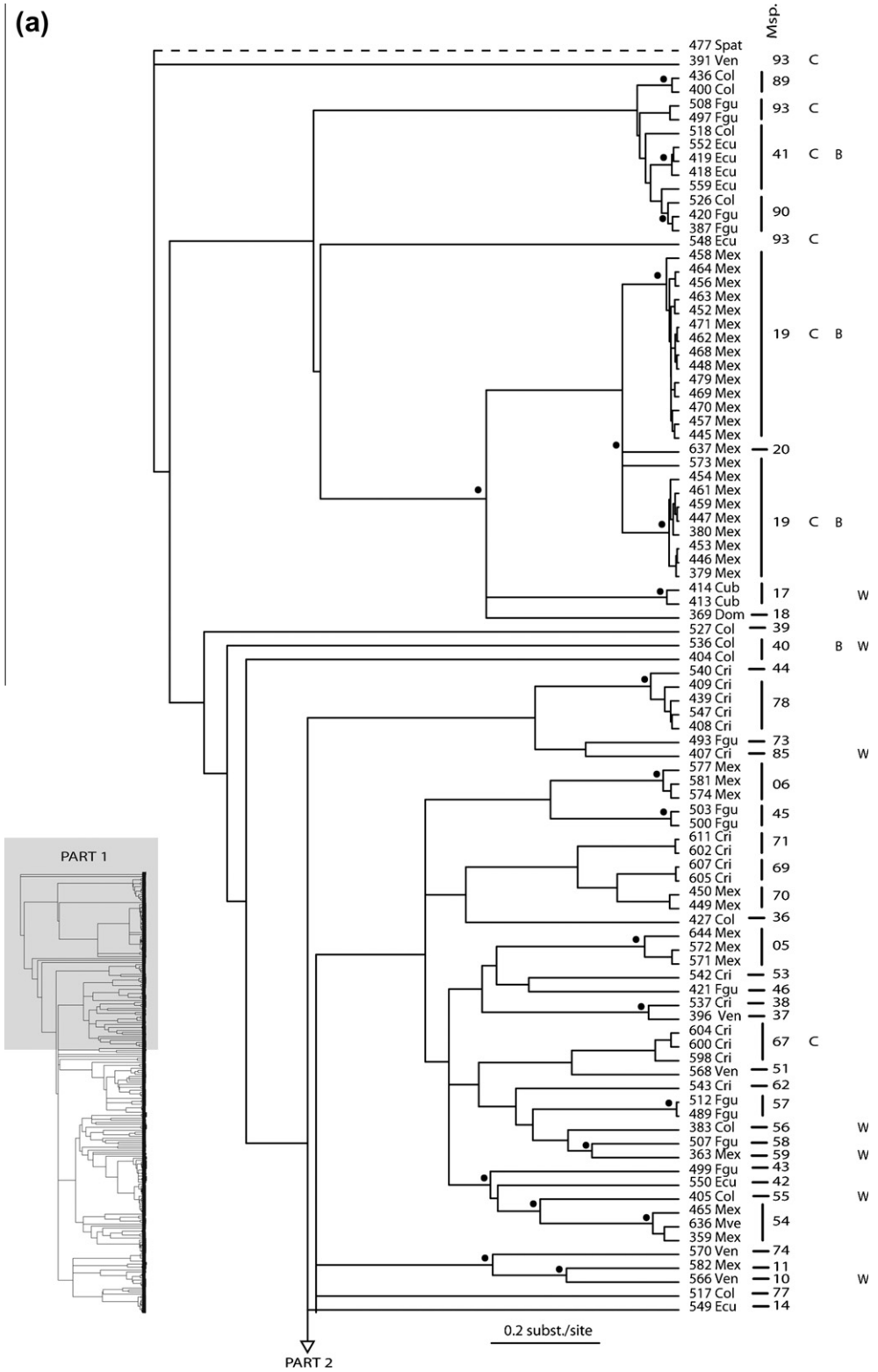

Fig. 1. Bayesian ultrametric tree reconstructed with the $both_{all}$ dataset using MrBayes and Pathd8 summarising the congruence/incongruence found between the morphospecies identified and the GMYC species recovered by COI_{all}, cyt b_{all} and *wg* using the selected sets of parameters and methods. Names of terminal taxa include their voucher numbers and the three-letter-code referring to the country where the specimen was collected (Bra = Brazil, Col = Colombia, Cri = Costa Rica, Cub = Cuba, Dom = Dominican Republic, Ecu = Ecuador, Fgu = French Guyana, Mex = Mexico, Tri = Trinidad and Tobago, Ven = Venezuela). Numbers to the right of terminal taxa indicate their assigned morphospecies (Msp.). Horizontal lines indicate singletons, vertical lines (or no lines) join individuals assigned to same morphospecies. Letters C, B and W to the right of morphospecies numbers indicate their incongruence with the GMYC species recovered with the COI, cyt *b* and *wg* ultrametric trees reconstructed using BEAST-l-c, BEAST-l-y and MrBayes + Pathd8, respectively. Black circles above/below branches represent Bayesian posterior probability values $\geqslant 0.95$.

Please cite this article in press as: Ceccarelli, F.S., et al. Species identification in the taxonomically neglected, highly diverse, neotropical parasitoid wasp genus *Notiospathius* (Braconidae: Doryctinae) based on an integrative molecular and morphological approach. Mol. Phylogenet. Evol. (2011), doi:10.1016/j.ympev.2011.10.018

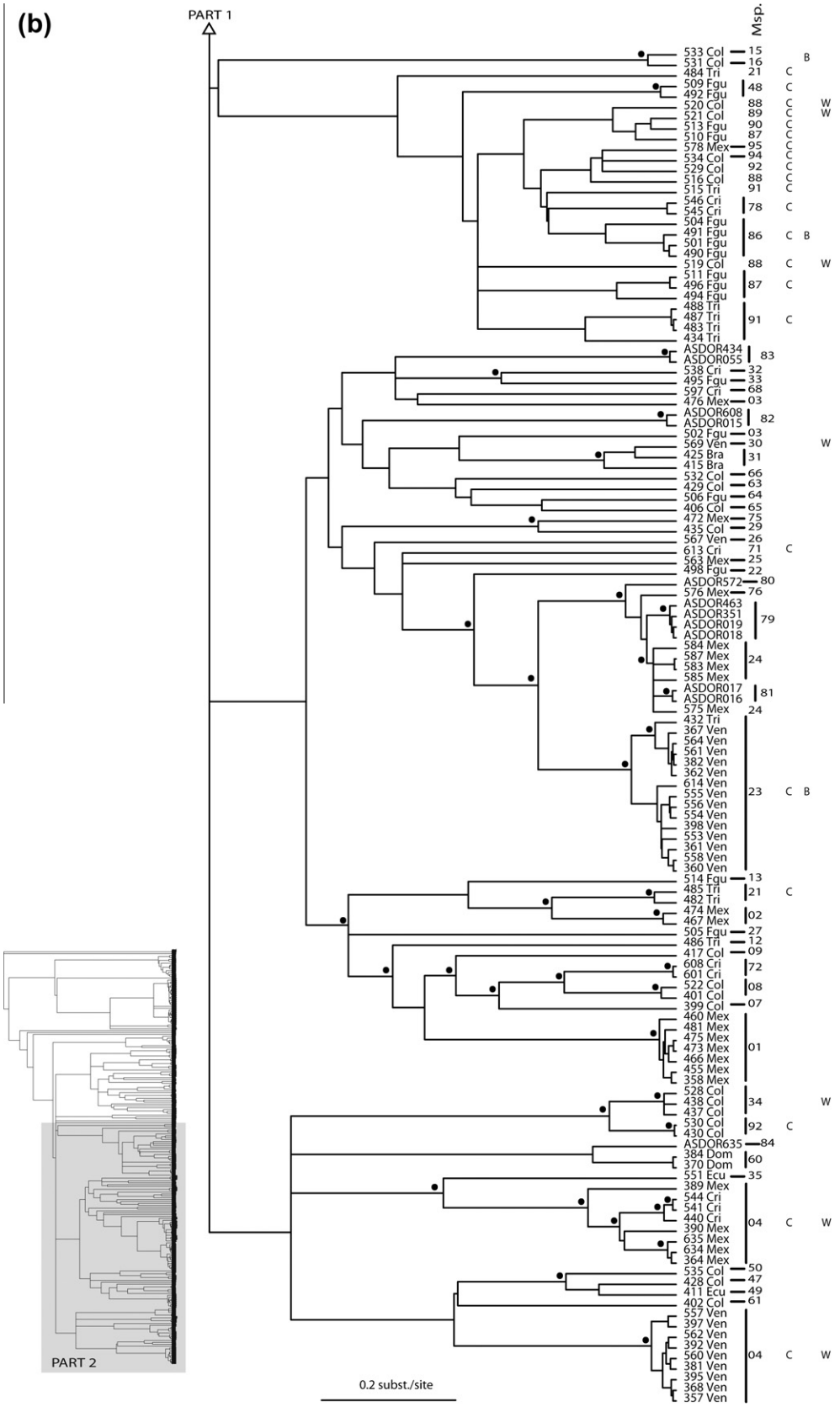

Fig. 1 (*continued*)

polyphyletic gene trees using nuclear loci relative to mt loci due to their larger effective population size (Funk and Omland, 2003; Hudson and Turelli, 2003), thus making the use of mt markers more reliable in reconstructing phylogenies of closely related

species. This and the lower amplification success of the *wg* gene fragment led us to exclude it for delimiting species in the group examined.

Based on our results, one can see that the performance of concatenated datasets in GMYC species delineation analyses could be limited by the degree of mismatch displayed by the markers included and by their different amounts of informative variation. Our concatenated mt dataset with missing data ($both_{all}$) had 44.1% mismatch and failed to recover some morphospecies whose specimens were only sequenced, and therefore only supported, either for COI_{all} or cyt b_{all}. Moreover, some clusters of specimens whose morphological congruence was concordant with cyt b_{all} but not by COI_{all} were not recovered by the concatenated datasets, which instead recovered the doubtful species clustering yielded by the latter marker. This could be because the nucleotidic variation of COI supporting the above incorrect species clusters was higher than the variation of cyt *b* that was congruent with morphology. It is commonly assumed that concatenating different gene markers, especially when they are linked as in the case of mt genes, reduces the stochastic effects of small datasets, and therefore improves the phylogenetic signal of the data. Some authors, however, have suggested that multigene phylogenetics can lead to incorrect phylogenetic inferences by giving more weight to a particular gene thus producing the wrong kind of average (Kubatko and Degnan, 2007; Degnan and Rosenberg, 2006). We therefore followed a conservative approach and delineated species based on the congruence observed among the markers examined separately (see below). Further studies employing various gene markers with different levels of mismatch will help to better understand the performance of concatenated datasets in GMYC analyses.

4.3. Species diversity in Notiospathius

Species boundaries studies based on integrative strategies that make use of both DNA sequence and morphological data, as well as more rigorous methods, are increasingly being conducted (e.g. Monaghan et al., 2009; Hawlitschek et al., 2011; Heethoff et al., 2011). In this study, we followed an integrative approach for rapidly identifying the number of species belonging to a large, understudied insect group. Our results have helped us identify a number of "stable" (*sensu* Padial and De la Riva, 2010) species, in the sense that they are supported by different lines of evidence that are not in conflict (morphology, at least one of the two mt markers and in some cases *wg*) and were subjected to a corroboration assessment through implementation of a coalescent-based method.

Based on the results obtained from the GMYC analysis of the separate mt datasets (COI_{all}, cyt b_{all}), a total of 71 species (67 *Notiospathius*, two *Tarasco*, and two *Masonius* spp.) can be unambiguously distinguished from the specimens examined. Moreover, we consider the seven GMYC species assigned to *Notiospathius* that were supported by both mt markers but not by morphology as potential cryptic species/pseudocryptic species, though this needs to be confirmed by additional nuclear sequence data (Smith et al., 2011). A list with the voucher numbers of the specimens belonging to our 71 "stable" species and the seven potential cryptic species can be found in the Supplementary material (S3). Only two of these unambiguously delimited species are currently described. Several of these species will be described shortly using the morphological and molecular information obtained from this study (Zaldívar-Riverón et al., unpublished data).

We still expect to discover a considerably higher number of undescribed species, since our specimen sampling was geographically scattered and superficial. This remarkable species diversity in a Neotropical braconid genus appears to be only lower than that found in the microgastrines *Cotesia* and *Apanteles*, the braconine *Bracon*, the rogadine *Aleiodes* and the doryctine *Heterospilus* (Yu et al., 2005). These genera, however, are not restricted to the Neotropics. We also found support for the placement of *Hansonorum* as a junior synonym *Notiospathius* (De Jesús-Bonilla et al., 2011) based on our morphological examinations and the molecular evidence obtained. We discovered that some of our stable species show intraspecific variation in the presence/absence of a tubercle on the hind coxa (see Supplementary material S2), which was the only feature previously employed to distinguish these genera (Marsh, 2002). Moreover, our phylogenetic analyses, carried out separately with two fast evolving mt and one nuclear marker, though regarded with caution, all suggest that there are multiple gains and losses of the aforementioned morphological feature. Further molecular phylogenetic studies will give insight on the evolution of this group of parasitoid wasps.

Acknowledgments

We thank Hans Clebsch for providing material from Dominican Republic, Mexico and Venezuela, and Juan José Martínez, Vladmir De Jesús-Bonilla, Valeria Salinas-Ramos, Itzel Zamora-Vilchis, Andrés Martínez-Aquino and Ana Cecilia Rodríguez-Perez for helping during our collecting trips in Mexico; Laura Marquez for sequencing part of the samples; Andrés Reséndiz-Flores, Karen López, Sofía Fernández-Parriego and Ma. Cristina Mayorga-Martínez for mounting and labelling the entomological material; Susana Guzmán-Gómez for helping with the microscope photography; and Kevin Williams, Alfried Vogler and two anonymous referees for their help improving the manuscript. This work was supported by Grants given by the Ministerio de Ciencia e Innovación (CGL2010-15786; Spain), Comisión Nacional para el Uso y Conocimiento de la Biodiversidad (CONABIO-HB033: Mexico), and the Consejo Nacional de Ciencia y Tecnología (CONACyT-511; Mexico) to AZR, the US National Science Foundation to MJS (DEB 0205982), and by a postdoctoral fellowship given by the Universidad Nacional Autónoma de México (UNAM-DGAPA) to FSC.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2011.10.018.

References

Barraclough, T.G., 2010. Evolving entities: towards a unified framework for understanding diversity at the species and higher levels. Philos. T. R. Soc. B 365, 1801–1813.

Barraclough, T.G., Birky, C.W., Burt, A., 2003. Diversification in sexual and asexual organisms. Evolution 57, 2166–2172.

Belshaw, R., Quicke, D.L.J., 1997. A molecular phylogeny of the aphidiinae (Hymenoptera: Braconidae). Mol. Phylogenet. Evol. 7, 281–293.

Birky, C.W., Wolf, C., Maughan, H., Herbertson, L., Henry, E., 2005. Speciation and selection without sex. Hydrobiologia 546, 29–45.

Blaxter, M.L., 2004. The promise of a DNA taxonomy. Philos. Trans. R. Soc. Lond. Ser. B-Biol. Sci. 359, 669–679.

Blaxter, M., Mann, J., Chapman, T., Thomas, F., Whitton, C., Floyd, R., Abebe, E., 2005. Defining operational taxonomic units using DNA barcode data. Philos. Trans. R. Soc. B 360, 1935–1943.

Britton, T., Oxelman, B., Vinnersten, A., Bremer, K., 2002. Phylogenetic dating with confidence intervals using mean path lengths. Mol. Phylogenet. Evol. 24, 58–65.

Britton, T., Anderson, C.L., Jacquet, D., Lundqvist, S., Bremer, K., 2007. Estimating divergence times in large phylogenetic trees. Syst. Biol. 56, 741–752.

Brower, A.V.Z., DeSalle, R., 1998. Patterns of mitochondrial versus nuclear DNA sequence divergence among nymphalid butterflies: the utility of wingless as a source of characters for phylogenetic inference. Insect Mol. Biol. 7, 73–82.

Burns, J.M., Janzen, D.H., Hajibabaei, M., Hallwachs, W., Hebert, P.D.N., 2008a. DNA barcodes and cryptic species of skipper butterflies in the genus Perichares in Area de Conservacion Guanacaste, Costa Rica. Proc. Natl. Acad. Sci. USA 105, 6350–6355.

Burns, J.M., Janzen, D.H., Hajibabaei, M., Hallwachs, W., Hebert, P.D.N., 2008b. DNA barcodes of closely related (but morphologically and ecologically distinct)

species of skipper butterflies (Hesperiidae) can differ by only one to three nucleotides. J. Lepido. Soc. 61, 138–153.
De Jesús-Bonilla, V.S., Nunes, J.F., Penteado-Dias, A., Csösz, S., Zaldívar-Riverón, A., 2011. A new synonym of the Neotropical parasitoid wasp genus *Notiospathius* (Braconidae, Doryctinae), with redescription of two species and description of five new species from Brazil. Zookeys 122, 71–90.
Degnan, J.H., Rosenberg, N.A., 2006. Discordance of species trees with their most likely gene trees. PLoS Genet. 2 (5), 0762–0768.
Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7.
Drummond, A.J., Suchard, M.A., 2010. Bayesian random local clocks, or one rate to rule them all. BMC Evol. Biol. 8.
Drummond, A.J., Nicholls, G.K., Rodrigo, A.G., Solomon, W., 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics 161, 1307–1320.
Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS. Biol. 4, 699–710.
Fontaneto, D., Herniou, E.A., Boschetti, C., Caprioli, M., Melone, G., Ricci, C., Barraclough, T.G., 2007. Independently evolving species in asexual bdelloid rotifers. PLoS. Biol. 5, 914–921.
Funk, D.J., Omland, K.E., 2003. Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annu. Rev. Ecol. Evol. Syst. 34, 397–423.
Gaston, K.J., 2000. Global patterns in biodiversity. Nature 405, 220–227.
Gattolliat, J.L., Monaghan, M.T., 2010. DNA-based association of adults and larvae in Baetidae (Ephemeroptera) with the description of a new genus Adnoptilum in Madagascar. J. N. Am. Benthol. Soc. 29, 1042–1057.
Hajibabaei, M., Singer, G.A.C., Hebert, P.D.N., Hickey, D.A., 2007. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. Trends Genet. 23, 167–172.
Hawlitschek, O., Porch, N., Hendrich, L., Balke, M., 2011. Ecological niche modelling and *n*DNA sequencing support a new, morphologically cryptic beetle species unveiled by DNA barcoding. PLoS One 6.
Hebert, P.D.N., Cywinska, A., Ball, S.L., DeWaard, J.R., 2003a. Biological identifications through DNA barcodes. Proc. R. Soc. Lond. Ser. B-Biol. Sci. 270, 313–321.
Hebert, P.D.N., Ratnasingham, S., deWaard, J.R., 2003b. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc. R. Soc. Lond. Ser. B-Biol. Sci. 270, S96–S99.
Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H., Hallwachs, W., 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. Proc. Natl. Acad. Sci. USA 101, 14812–14817.
Heethoff, M., Laumann, M., Weigmann, G., Raspotnig, G., 2011. Integrative taxonomy: combining morphological, molecular and chemical data for species delineation in the parthenogenetic *Thypochthonius tectorum* complex (Acari, Oribatida, Trhypochthoniidae). Front. Zool. 8, 2.
Hong, S.H., Bunge, J., Jeon, S.O., Epstein, S.S., 2006. Predicting microbial species richness. Proc. Natl Acad. Sci. USA 103, 117–122.
Hudson, R.R., Turelli, M., 2003. Stochasiticity overrules the "three-times rule": genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. Evolution 57, 182–190.
Janzen, D.H., Hallwachs, W., Blandin, P., Burns, J.M., Cadiou, J.M., Chacon, I., Dapkey, T., Deans, A.R., Epstein, M.E., Espinoza, B., Franclemont, J.G., Haber, W.A., Hajibabaei, M., Hall, J.P.W., Hebert, P.D.N., Gauld, I.D., Harvey, D.J., Hausmann, A., Kitching, I.J., Lafontaine, D., Landry, J.F., Lemaire, C., Miller, J.Y., Miller, J.S., Miller, L., Miller, S.E., Montero, J., Munroe, E., Green, S.R., Ratnasingham, S., Rawlins, J.E., Robbins, R.K., Rodriguez, J.J., Rougerie, R., Sharkey, M.J., Smith, M.A., Solis, M.A., Sullivan, J.B., Thiaucourt, P., Wahl, D.B., Weller, S.J., Whitfield, J.B., Willmott, K.R., Wood, D.M., Woodley, N.E., Wilson, J.J., 2009. Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. Mol. Ecol. Resour. 9, 1–26.
Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16, 111–120.
Kubatko, L.S., Degnan, J.H., 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56 (1), 17–24.
Lahaye, R., Van der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., Maurin, O., Duthoit, S., Barraclough, T.G., Savolainen, V., 2008. DNA barcoding the floras of biodiversity hotspots. Proc. Natl. Acad. Sci. USA 105, 2923–2928.
Lanave, C., Preparata, G., Saccone, C., Serio, G., 1984. A new method for calculating evolutionary substitution rates. J. Mol. Evol. 20, 86–93.
Lin, C.P., Danforth, B.N., 2004. How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets. Mol. Phylogenet. Evol. 30, 686–702.
Maddison, D.R., Maddison, W.P., 2000. MacClade 4: Analysis of Phylogeny and Character Evolution. Sinauer Associates, Sunderland, Massachusetts.
Marsh, P.M., 2002. The Doryctinae of Costa Rica (excluding the genus *Heterospilus*). Mem. Am. Entomol. Inst. 70, 1–319.
Matthews, R.W., Marsh, P.M., 1973. *Notiospathius*, a new neotropical genus (Hymenoptera: Braconidae). J. Wash. Acad. Sci. 63, 73–75.
McFadden, C.S., Benayahu, Y., Pante, E., Thoma, J.N., Nevarez, P.A., France, S.C., 2011. Limitations of mitochondrial gene barcoding in Octocorallia. Mol. Ecol. Resour. 11, 19–31.
Meyer, C.P., Paulay, G., 2005. DNA barcoding: error rates based on comprehensive sampling. PloS Biol. 3, 2229–2238.
Miller, S.E., 2007. DNA barcoding and the renaissance of taxonomy. Proc. Nat. Acad. Sci. USA 104, 4775–4776.
Miller, K.B., Alarie, Y., Wolfe, G.W., Whiting, M.F., 2005. Association of insect life stages using DNA sequences: the larvae of Philodytes umbrinus (Motschulsky) (Coleoptera: Dytiscidae). Syst. Entomol. 30, 499–509.
Monaghan, M.T., Balke, M., Gregory, T.R., Vogler, A.P., 2005. DNA-based species delineation in tropical beetles using mitochondrial and nuclear markers. Philos. Trans. R. Soc. B-Biol. Sci. 360, 1925–1933.
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 delineation. Syst. Biol. 58, 298–311.
Nylander, J.A.A., 2004. MrModeltest v2. Program Distributed by the Author. Evolutionary Biology Centre, Uppsala University, Sweden.
Padial, J.M., De la Riva, I., 2010. A response to recent proposals for integrative taxonomy. Biol. J. Linn. Soc. 101, 747–756.
Papadopoulou, A., Anastasiou, I., Vogler, A.P., 2010. Revisiting the insect mitochondrial molecular clock: the mid-aegean trench calibration. Mol. Biol. Evol. 27, 1659–1672.
Pauls, S.U., Blahnik, R.J., Zhou, X., Wardwell, C.T., Holzenthal, R.W., 2010. DNA barcode data confirm new species and reveal cryptic diversity in Chilean Smicridea (Smicridea) (Trichoptera: Hydropsychidae). J. N. Am. Benthol. Soc. 29, 1058–1074.
Pinzon-Navarro, S., Barrios, H., Murria, C., Lyal, C.H.C., Vogler, A.P., 2010. DNA-based taxonomy of larval stages reveals huge unknown species diversity in Neotropical seed weevils (genus Conotrachelus): relevance to evolutionary ecology. Mol. Phylogenet. Evol. 56, 281–293.
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. Syst. Biol. 55, 595–609.
R Core Development Team, 2009. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.
Sáez, A.G., Lozano, E., 2005. Body doubles. Nature 433, 111.
Sanderson, M.J., 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol. Biol. Evol. 19, 101–109.
Sanderson, M.J., 2003. R8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics 19, 301–302.
Sheffield, C.S., Hebert, P.D.N., Kevan, P.G., Packer, L., 2009. DNA barcoding a regional bee (Hymenoptera: Apoidea) fauna and its potential for ecological studies. Mol. Ecol. Resour. 9, 196–207.
Simmons, R.B., Weller, S.J., 2001. Utility and evolution of cytochrome b in insects. Mol. Phylogenet. Evol. 20, 196–210.
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. Ann. Entomol. Soc. Am. 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). Proc. Natl. Acad. Sci. USA 103, 3657–3662.
Smith, M.A., Rodriguez, J.J., Whitfield, J.B., Deans, A.R., Janzen, D.H., Hallwachs, W., Hebert, P.D.N., 2008. Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. Proc. Natl. Acad. Sci. USA 105, 12359–12364.
Smith, M.A., Fernandez-Triana, J., Roughley, R., Hebert, P.D.N., 2009. DNA barcode accumulation curves for understudied taxa and areas. Mol. Ecol. Resour. 9, 208–216.
Smith, M.A., Eveleigh, E.S., McCann, K.S., Merilo, M.T., McCarthy, P.C., Van Rooyen, K.I., 2011. Barcoding a quantified food web: crypsis, concepts, ecology and hypotheses. PLOS One 6, e14424.
Swofford, D., 2001. PAUP*: Phylogenetic Analysis Using Parsimony (*and other methods). Sinauer Associates Inc., Sunderlands, MA.
Thorne, J.L., Kishino, H., 2002. Divergence time and evolutionary rate estimation with multilocus data. Syst. Biol. 51, 689–702.
Ward, R.D., Zemlak, T.S., Innes, B.H., Last, P.R., Hebert, P.D.N., 2005. DNA barcoding Australia's fish species. Philos. Trans. R. Soc. B-Biol. Sci. 360, 1847–1857.
Wertheim, J.O., Sanderson, M.J., Worobey, M., Bjork, A., 2010. Relaxed molecular clocks, the bias-variance trade-off, and the quality of phylogenetic inference. Syst. Biol. 59, 1–8.
Yang, Z.H., Rannala, B., 2010. Bayesian species delimitation using multilocus sequence data. Proc. Natl. Acad. Sci. USA 107, 9264–9269.
Yu, D., van Achterberg, K., Horstmann, K., 2005. World Ichneumonoidea. 2004. Taxonomy, Biology, Morphology and Distribution. Taxapad, Vancouver, Canada.
Zaldívar-Riverón, A., De Jesus-Bonilla, V.S., 2010. Redescription of species of the Neotropical parasitoid Notiospathius Mathews et Marsh (Braconidae: Doryctinae) based on their nineteenth and early twentieth century types. Zootaxa, 31–42.
Zaldívar-Riverón, A., Belokobylskij, S.A., Leon-Regagnon, V., Martinez, J.J., Briceno, R., Quicke, D.L.J., 2007. A single origin of gall association in a group of parasitic wasps with disparate morphologies. Mol. Phylogenet. Evol. 44, 981–992.
Zaldívar-Riverón, A., Belokobylskij, S.A., Leon-Regagnon, V., Briceno-G, R., Quicke, D.L.J., 2008. Molecular phylogeny and historical biogeography of the cosmopolitan parasitic wasp subfamily Doryctinae (Hymenoptera: Braconidae). Invert. Syst. 22, 345–363.

Please cite this article in press as: Ceccarelli, F.S., et al. Species identification in the taxonomically neglected, highly diverse, neotropical parasitoid wasp genus *Notiospathius* (Braconidae: Doryctinae) based on an integrative molecular and morphological approach. Mol. Phylogenet. Evol. (2011), doi:10.1016/j.ympev.2011.10.018

Zaldívar-Riverón, A., Martinez, J.J., Ceccarelli, F.S., De Jesus-Bonilla, V.S., Rodriguez-Perez, A.C., Resendiz-Flores, A., Smith, M.A., 2010. DNA barcoding a highly diverse group of parasitoid wasps (Braconidae: Doryctinae) from a Mexican nature reserve. Mitochon. DNA 21, 18–23.

Please cite this article in press as: Ceccarelli, F.S., et al. Species identification in the taxonomically neglected, highly diverse, neotropical parasitoid wasp genus *Notiospathius* (Braconidae: Doryctinae) based on an integrative molecular and morphological approach. Mol. Phylogenet. Evol. (2011), doi:10.1016/j.ympev.2011.10.018