MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2015) **15**, 268–277 doi: 10.1111/1755-0998.12304

DNA barcoding gap: reliable species identification over morphological and geographical scales

KLEMEN ČANDEK* and MATJAŽ KUNTNER*†‡
*Institute of Biology, Scientific Research Centre of the Slovenian Academy of Sciences and Arts, Novi Trg 2, 1000 Ljubljana, Slovenia, †Centre for Behavioural Ecology and Evolution, College of Life Sciences, Hubei University, 368 Youyi Road, 430062 Wuhan, China, ‡Department of Entomology, National Museum of Natural History, Smithsonian Institution, PO Box 37012, Washington, DC 20013-7012, USA

Abstract

The philosophical basis and utility of DNA barcoding have been a subject of numerous debates. While most literature embraces it, some studies continue to question its use in dipterans, butterflies and marine gastropods. Here, we explore the utility of DNA barcoding in identifying spider species that vary in taxonomic affiliation, morphological diagnosibility and geographic distribution. Our first test searched for a 'barcoding gap' by comparing intra- and interspecific means, medians and overlap in more than 75 000 computed Kimura 2-parameter (K2P) genetic distances in three families. Our second test compared K2P distances of congeneric species with high vs. low morphological distinctness in 20 genera of 11 families. Our third test explored the effect of enlarging geographical sampling area at a continental scale on genetic variability in DNA barcodes within 20 species of nine families. Our results generally point towards a high utility of DNA barcodes in identifying spider species. However, the size of the barcoding gap strongly depends on taxonomic groups and practices. It is becoming critical to define the barcoding gap statistically more consistently and to document its variation over taxonomic scales. Our results support models of independent patterns of morphological and molecular evolution by showing that DNA barcodes are effective in species identification regardless of their morphological diagnosibility. We also show that DNA barcodes represent an effective tool for identifying spider species over geographic scales, yet their variation contains useful biogeographic information.

Keywords: barcoding gap, biogeography, CO1, DNA barcodes, morphology, spiders

Received 10 May 2014; revision received 6 July 2014; accepted 16 July 2014

Introduction

Modern taxonomy increasingly relies on molecular tools, the most popular being DNA barcoding, a straightforward and relatively cheap method for species identification. Ever since Hebert and colleagues proposed a roughly 650-nucleotide-long segment of cytochrome c oxidase subunit 1 (CO1) gene as a 'DNA barcode' tool in animals (Hebert *et al.* 2003), the method has been a subject of infinite debates about its reliability and usefulness (reviewed in Taylor & Harris 2012). While most literature embraces it (Hebert *et al.* 2003, 2004b; Barrett & Hebert 2005; Hajibabaei *et al.* 2006) or shows its superiority over the use of mitochondrial markers other than CO1 (Aliabadian *et al.* 2009), some papers specific to its utility for certain taxa continue to question DNA barcoding as useful, for example in dipterans, butterflies and marine gastropods (Meyer & Paulay 2005; Meier *et al.* 2006; Wiemers & Fiedler 2007).

In this study, we test the utility of DNA barcodes for spider species identification. Spiders are hyperdiverse invertebrates with more than 44 000 described (Platnick 2014), and over 100 000 expected species (Agnarsson *et al.* 2013). They are found in most terrestrial habitats, but spiders of different lineages vary considerably in biology, ecology and dispersal abilities, and consequently also in their geographic distributions (Bell *et al.* 2005; Kuntner & Agnarsson 2011b). Many species possess very distinct behaviour and morphological characteristics, and their sizes range from 0.43 to 280 mm (Smith 2008; Foelix 2010). These components of diversity may suggest substantial difficulties in applying one single identification method to all spiders and expecting its reliability. Despite the promise of DNA barcoding, only a handful of studies have used it in spider research (Barrett & Hebert 2005; Hebert & Barrett 2005; Prendini 2005; Arnedo & Ferrández 2007; Longhorn *et al.* 2007; Blagoev

Correspondence: Klemen Čandek and Matjaž Kuntner,
Fax: +386 1 425 7797;
E-mails: klemen.candek@gmail.com;
kuntner@gmail.com

© 2014 John Wiley & Sons Ltd

et al. 2009; Kuntner & Agnarsson 2011a; Hendrixson *et al.* 2013). However, none of these studies rigorously tested the barcoding utility across morphological and geographical scales in spiders, as we do here.

Species identification with DNA barcodes is only reliable if a significant difference between the average intraspecific and the average interspecific genetic distance can be consistently detected (Hebert *et al.* 2003, 2004b; Barrett & Hebert 2005); Hebert *et al.* defined such 'barcoding gap' as the existence of at least 10 times greater average interspecific distance over the average intraspecific genetic distance (Hebert *et al.* 2004b). Early studies on birds and arthropods, including spiders, supported the gap's existence (Hebert *et al.* 2004b; Barrett & Hebert 2005; Hajibabaei *et al.* 2006) while other studies did not (Meyer & Paulay 2005; Meier *et al.* 2006; Wiemers & Fiedler 2007). Among its critics, Meyer & Paulay (2005) suggested that the 'barcoding gap' is an artefact of insufficient sampling across taxa and individuals. While numerous studies apply methods other than barcoding gap analysis to delimit species (Knowles & Carstens 2007; Rosenberg 2007; Cummings *et al.* 2008; Rodrigo *et al.* 2008; Bertolazzi *et al.* 2009; O'Meara 2010; Yang & Rannala 2010; Zaldívar-Riverón *et al.* 2010; Masters *et al.* 2011; Zhang *et al.* 2011; Boykin *et al.* 2012; Fujita *et al.* 2012; Nuñez *et al.* 2012; Vuataz *et al.* 2012; Weitschek *et al.* 2013; White *et al.* 2014), we here only focus on the classic barcoding gap in order for our results to be comparable with studies on other organisms. Thus, our first goal was to rigorously test for the existence of a barcoding gap in spiders. If the gap did exist, we would detect a significantly greater average interspecific distance compared with the average intraspecific distance in several independent spider lineages.

The concept of a molecular clock (Zuckerkandl & Pauling 1965) postulates that mutations emerge, fixate and accumulate predictably through time, which allows estimation of phyletic and speciation time (Janecka *et al.* 2012). On the other hand, morphological evolution runs less predictably. A logical consequence of this disparity is that morphological and molecular evolution ought to be disconnected (Bromham *et al.* 2002; Davies & Savolainen 2006; Goldie *et al.* 2011). If so, this would imply that DNA barcode utility is difficult to predict from morphological taxonomic variation. For example, at one extreme, a group of ecologically and morphologically distinct butterflies varied in only 1–3 nucleotides (Burns *et al.* 2007), while at the other extreme, Hebert *et al.* (2004a) redefined 10 (cryptic) species that could previously not have been diagnosed morphologically. Because similar disparities between molecules and morphology are a norm rather than exception (Meier *et al.* 2006; BioEssays 2009; Tavares *et al.* 2011), the most reliable species identification would be through combination of morphology and DNA barcodes (Delsinne *et al.* 2012), or even with the addition of other types of data such as geography, behaviour, ecology, etc. (Yassin *et al.* 2010). Our second goal was to test whether barcoding works as reliably in morphologically similar vs. distinct congeneric spider species. If DNA barcodes were reliable, we would detect no correlation between species morphological characteristics and interspecific molecular divergences.

DNA barcoding studies not only propose their utility in discovering species in wide-ranging taxa (Johnsen *et al.* 2010; Nijman & Aliabadian 2013) but also point to their biogeographic and phylogeographic utility (Carr *et al.* 2011; Nwani *et al.* 2011; Webster *et al.* 2012; Ashfaq *et al.* 2014). Theory predicts increased variation in genetic distance over increased geographical scale (Wright 1943; Nekola & White 1999), and empirical studies confirm such intraspecific trends (Avise 2000). A question arises whether DNA barcodes can reliably identify taxa spread over large geographic areas. On the one hand, Negri *et al.* (2012) used DNA barcodes over large geographic scales (North to South America) to delimit previously conspecific ant taxa, but on the other, a study on water beetles (Bergsten *et al.* 2012) interpreted barcodes to be less useful for identifying species with increased geographic ranges. While certain studies suggested that geography and genetic distances were not codependent (Hebert *et al.* 2004b, 2010), Bergsten *et al.* (2012) attributed their deviations from standard theories to flawed testing on small geographic scales and on organisms with exceptional dispersal abilities. Therefore, it remains quite possible that over large geographic scales, increased intraspecific and decreased interspecific distances render identification with DNA barcodes less effective. Our third goal was to test this hypothesis in spiders. We predicted that the increase of intraspecific distances with increased sampling from one continent to two would limit successful species identification with DNA barcodes.

Materials and methods

Data acquisition

We combined our original COI sequences with others downloaded from BOLD systems (Ratnasingham & Hebert 2007) (Appendices 1–3). In BOLD systems, we targeted all public COI sequences of 600–700 bp on 15 January 2014, and retained those with unequivocal species names from the two orbweaving families (Araneidae, Tetragnathidae) and the cursorial wolf spiders (Lycosidae). Furthermore, we targeted specific 20 genera with whose species identification we were familiar for morphological

© 2014 John Wiley & Sons Ltd

analysis and specific 20 species with intercontinental distribution for biogeographic analysis.

Genetic distances

We aligned the sequences using CLUSTALW in MEGA 5.1 (Tamura *et al.* 2011) and computed genetic distances using Kimura 2-parameter (K2P) (Kimura 1980). Although some authors question K2P model as the most appropriate metric (Srivathsan & Meier 2012) or suggest using simpler statistics such as *p* distances (Collins *et al.* 2012), we use K2P because it represents the standard in DNA barcoding literature and therefore facilitates comparisons.

Testing the barcoding gap

To test our fist hypothesis, we first trimmed the distance data set for the sequences with the highest 5% intraspecific distances and the 5% lowest interspecific distances, thereby removing the most likely misidentifications (Meier *et al.* 2006, 2008). Because morphological misidentifications are common in the literature and particularly rampant in public databases, reaching well over 5% (Oliver & Beattie 1996; Scott & Hallam 2002; Bridge *et al.* 2003; Haase *et al.* 2006; Dexter *et al.* 2010; Hull *et al.* 2010; Shea *et al.* 2011; Conn *et al.* 2013), we believe that eliminating the 5% of the most likely errors is a statistically justifiable approach and a conservative test. We looked for overlap between intra- and interspecific K2P distances for each separate family and for all three families combined, and for statistically significant differences between intra- and interspecific K2P distances. We also checked for the classical barcoding gap by verifying a tenfold mean K2P distances difference. Although most barcoding literature reports the differences in means (Hebert *et al.* 2004b), we additionally report the differences in medians in those cases where the data were not normally distributed.

DNA barcoding and morphology

To test our second hypothesis, we selected 20 genera with whose species identification we were familiar. According to our experience, we assigned the genera to one of the two categories and tested statistical differences in K2P distances between the groups: (i) 'high' grouped those genera whose species identification was straightforward due to high morphological distinctness in at least one sex; and (ii) 'low' included those genera whose species were morphologically very similar and therefore difficult to identify. The genera were represented with at least five species. We first averaged the calculated K2P distances in each genus, then compared the two established groups with parametric statistic tests.

DNA barcoding and biogeography

To test our third hypothesis, we selected 20 species distributed in both North America and Europe that were represented in our data set with at least seven individuals per continent. We removed all sequences that lacked the country of origin, and randomly selected 10 individuals of species with good specimen representation. Depending on their origin, we assigned all selected individuals to one of the two categories, North America (NA) and Europe (E). We then calculated intraspecific K2P distances within each group (NA and E) and between the groups (category NAE).

Statistical analyses

We used Kolmogorov–Smirnov test of normality and parametric and nonparametric statistics in SPSS.

Results

We operated with 1203 DNA barcodes for the barcoding gap analysis (Appendix S1, supporting information), 1633 DNA barcodes for the analysis of DNA barcode utility over morphological scales (Appendix S2, supporting information) and 382 for the analysis of their utility over geographical scales (Appendix S3, supporting information).

Barcoding gap

Our comparison of 1203 individual barcodes belonging to 162 species and three spider families (Appendix S1; Table 1; Fig. 1) found nonnormally distributed K2P data (Kolmogorov–Smirnov test for all groups $P < 0.001$), and statistically significant differences in K2P values between all intraspecific and interspecific comparisons within families (Mann–Whitney test for: Araneidae $Z = -83.98$, $P < 0.001$; Lycosidae $Z = -134.41$, $P < 0.001$; Tetragnathidae $Z = -94.99$, $P < 0.001$) and when the families were combined (Mann–Whitney for All $Z = -186.78$, $P < 0.001$).

In Araneidae, we found a substantial overlap between the intra- and the interspecific K2P distances (Fig. 2) with an eightfold difference in means; the intraspecific mean was 0.011 ($N = 4145$), and the interspecific mean was 0.088 ($N = 7339$). However, because the data were not normally distributed, a statistically precise measure is to report the medians. In Araneidae, the interspecific median was 24 times greater than the intraspecific median (intraspecific median 0.003 vs. interspecific median 0.072). The overlap between the largest intraspecific and the smallest interspecific K2P distances disappeared at the 90th intraspecific (K2P = 0.028) and the 10th

© 2014 John Wiley & Sons Ltd

Table 1 Descriptive statistics for intraspecific and interspecific K2P (Kimura 2-parameter) distances, number of individuals, species and comparisons for Araneidae, Lycosidae, Tetragnathidae and all three families combined

		Intraspec. stat.	Interspec. stat.	No. of individuals/species	No. of inter-/intra comparisons
Araneidae	Mean	0.01115	0.08795	399/57	4145/7339
	SE	0.000274	0.000487		
	Median	0.003	0.072		
	SD	0.017615	0.041726		
	Minimum	0	0.041		
	Maximum	0.067	0.22		
	Interquartile range	0.014	0.068		
Lycosidae	Mean	0.00327	0.06019	578/79	6960/44331
	SE	0.000045	0.000061		
	Median	0.002	0.059		
	SD	0.003787	0.012815		
	Minimum	0	0.036		
	Maximum	0.012	0.164		
	Interquartile range	0.006	0.016		
Tetragnathidae	Mean	0.01507	0.17572	226/26	4539/8909
	SE	0.00021	0.000206		
	Median	0.012	0.173		
	SD	0.014132	0.019403		
	Minimum	0	0.131		
	Maximum	0.091	0.272		
	Interquartile range	0.013	0.029		
All	Mean	0.00878	0.08054	1203/162	15644/60579
	SE	< 0.00001	0.000183		
	Median	0.005	0.062		
	SD	0.013154	0.045041		
	Minimum	0	0.036		
	Maximum	0.091	0.272		
	Interquartile range	0.011	0.025		

interspecific percentile (K2P = 0.044). These K2P values therefore represent the thresholds for correct identification of 90% of individuals.

In the remaining two families, there was no overlap between the intra- and the interspecific K2P distances (Fig. 2), and such gap alone suggests 100% identification accuracy. In Lycosidae, the differences between the intra- and the interspecific means and medians were 20 and 29.5 times, respectively; the intraspecific mean was 0.003 and the median was 0.002 (N = 6960), the interspecific mean was 0.06 and the interspecific median 0.059 (N = 44 331). The lycosid lowest and highest thresholds were at 0.012 and 0.036 for intra- and interspecific K2P distances, respectively. In Tetragnathidae, the differences between the intra- and the interspecific means and medians were 11.7 and 14.4 times, respectively; the intraspecific mean was 0.015 and the median was 0.012 (N = 4539), the interspecific mean was 0.176 and the interspecific median 0.173 (N = 8909). The tetragnathid lowest and highest thresholds were at 0.091 and 0.131 for intra- and interspecific K2P distances, respectively.

When combining the data from all three families (Fig. 2), the differences in K2P intra- and interspecific means and medians were 9 and 12.4 times, respectively; the intraspecific mean was 0.009 and the median was 0.005 (N = 15 644), the interspecific mean was 0.081 and the interspecific median 0.062 (N = 60 579). The detected overlap between the highest intraspecific and the lowest interspecific K2P distances disappeared at the 95th intraspecific (K2P = 0.036) and the 5th interspecific percentile (K2P = 0.044), suggesting a 95% species identification success.

DNA barcoding and morphology

As the data (for selected species and individuals, see Appendix S2) were normally distributed (Kolmogorov–Smirnov: low P = 0.2, high P = 0.082), we employed a one-way ANOVA that, in support of our second hypothesis, showed no statistically significant difference (F = 0.134 P = 0.719) between the groups (Fig. 3; 'low' mean = 0.103 (N = 70 859), 'high' mean = 0.081 (N = 30 560)).

DNA barcoding and biogeography

The three groups contained calculated intraspecific K2P distances between individuals from North America only

© 2014 John Wiley & Sons Ltd

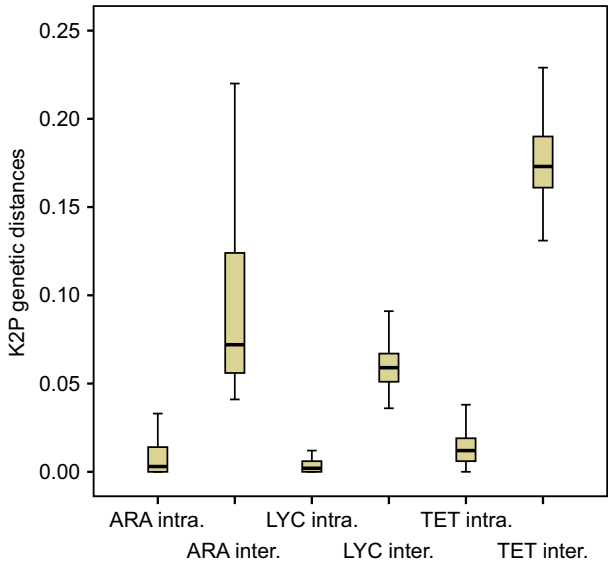

Fig. 1 Significantly different genetic distances between and within species in three spider families. Box plots represent intraspecific and interspecific genetic distances calculated using Kimura 2-parameter (K2P) model for each family (ARA=Araneidae, LYC=Lycosidae, TET=Tetragnathidae, intra.=intraspecific, inter.=interspecific), and error bars are interquartile ranges.

(group NA, mean = 0.0059, median = 0.003, $N = 742$), from Europe only (group E, mean = 0.0051, median = 0.0022, $N = 742$), and between individuals from North America and Europe (group NAE, mean = 0.0101, median = 0.0092, $N = 1665$) (Fig. 4; for selected species and individuals, see Appendix S3). The data in all three groups were not normally distributed (Kolmogorov–Smirnov $P < 0.001$). We found statistically significant differences between the groups NA and NAE (Mann–Whitney, $Z = -12.27$, $P < 0.001$) and between the groups E and NAE (Mann–Whitney, $Z = -14.71$, $P < 0.001$). The average K2P distances calculated between representatives from different continents (group NAE) were higher compared with those calculated from the representatives from a single continent (NA and E; see Fig. 4). The NAE mean was 1.7 and two times greater, and its median was 3 and 4.5 times greater than that of NA and E, respectively (for other statistics, see Table 2).

Discussion

Barcoding gap

Whether or not we confirmed the existence of the barcoding gap depends on its definition. Strictly following the original definition (Hebert *et al.* 2004b), we can confirm the barcoding gap in the tested data sets of the two families (Tetragnathidae and Lycosidae) where the mean interspecific K2P distances exceeded the mean intraspecific ones over 10 times (specifically, 11.4 and 20 times). However, in the case of our three-family data set, the appropriate average metric is the median, not the mean, due to data distribution (Links *et al.* 2012; Van Der Bank *et al.* 2013). Using the median, we confirmed the barcoding gap in all three families and in the combined data set, as the differences in the medians were always greater than 10 times (Araneidae 24, Lycosidae 29.5, Tetragnathidae 14.4 and All 12.4 times), and all intraspecific groups had significantly lower K2P distances compared with interspecific groups. A statistically accurate approach therefore established the barcoding gap in all the groups.

Another approach to detect the barcoding gap is to check the overlap between the lowest interspecific and the highest intraspecific genetic distances (Meier *et al.* 2008). The fact that we found no such overlap in the families Lycosidae and Tetragnathidae may alone suggest that species identification using barcodes would be 100% successful in these data sets. On the other hand, the detected overlap in the family Araneidae suggests reduced barcode effectiveness in this group. The obvious question, then, is to quantify barcoding effectiveness. According to our crude estimate that excluded the data between the 90th intraspecific and 10th interspecific percentile and thus established the K2P thresholds at 0.028 (intraspecific) and 0.044 (interspecific), species identifications in Araneids would be 90% accurate.

The differences in the detected genetic distance patterns among these three families may be real (due to different biologies), but are more likely merely due to differing taxonomic practices. Directly supporting the latter are the differences in average genetic intraspecific distances (Tetragnathidae mean at 1.5% compared to Araneidae at 1.1% and Lycosidae at only 0.3%). In addition, all three families were not represented by equal genetic samples, the family Lycosidae having a larger number of comparisons (51 291 compared to 11 493 and 13 448), which biases the total average. Although the goal of our study was to test barcoding effectiveness in spiders in general, combining all genetic distance data from the three families (Fig. 2) may thus produce spurious results. Despite these difficulties, a combined familial comparison still found a barcoding gap through detecting over 10 times greater interspecific than intraspecific median K2P genetic distance. Despite a factual overlap between the intra- and interspecific distances, it disappears at the 95th intraspecific (K2P = 0.036) and the 5th interspecific percentile (K2P = 0.044), suggesting a 95% species identification success in all spiders.

It is becoming clear that the term *barcoding gap* needs a more precise and statistically sound definition. Our results leave little doubt that the gap is obvious in at least two families, the cursorial ground spiders of the family

© 2014 John Wiley & Sons Ltd

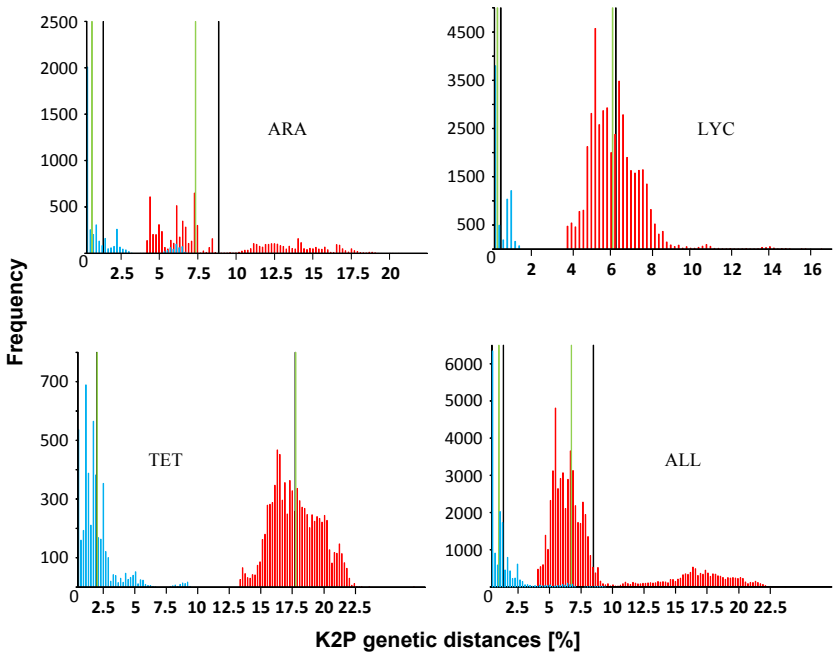

Fig. 2 A test of the barcoding gap in three spider families individually and combined. Frequency distributions of intraspecific and interspecific (congeneric) genetic divergences calculated using Kimura 2-parameter (K2P) model in the spider families Araneidae (ARA) with total number of 4145 intra- and 7339 interspecific comparisons across 57 species, Lycosidae (LYC), with total number of 6960 intra- and 44 331 interspecific comparisons across 79 species, Tetragnathidae (TET) with total number of 4539 intra- and 8909 interspecific comparisons across 26 species and for all three families combined (ALL) with total number of 15 644 intra- and 60 579 interspecific comparisons across 162 species.

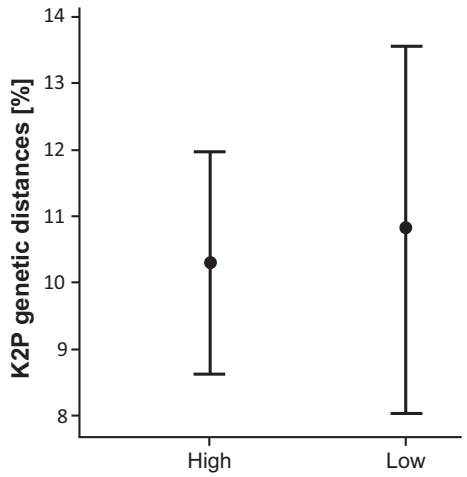

Fig. 3 Analysis of barcoding utility over morphological scales found no statistically significant differences in mean K2P values between the groups representing genera with 'high' and 'low' diagnosibility of their species. Error bars represent 95% confidence intervals.

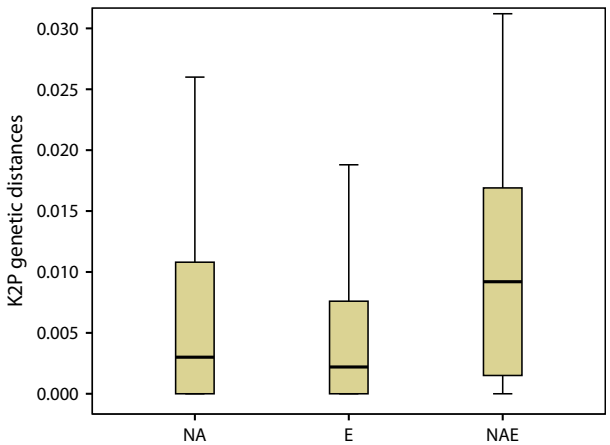

Fig. 4 Analysis of barcoding utility over geographical scales. Box plots represent medians and interquartile ranges for intraspecific K2P genetic distances of individuals from North America (NA), from Europe (E), and for intraspecific K2P distances computed between populations from North America and Europe (NAE).

Lycosidae and the aerial web builders of the family Tetragnathidae, and depending on the statistics used, also in the third tested family, the Araneidae. Judging from the data overlap, species identification using barcoding should be fully effective in the former two families, and 90% effective in the latter. Taking all groups combined, one should be able to identify approximately 95% of spider species. We conclude that DNA barcodes are informative for identifying spider species but that the size of the barcoding gap strongly depends on taxonomic groups and practices. We therefore concur with Yassin *et al.* (2010) that a taxonomically universal threshold in the barcoding gap is impossible to define.

DNA barcoding and morphology

As predicted, we found no significant differences in interspecific genetic distances between two groups

© 2014 John Wiley & Sons Ltd

Table 2 Descriptive statistics and number of intraspecific K2P comparisons between individuals of species only from North America, only from Europe and between populations from North America and Europe

		Descriptive statistics	No. of intraspec. comparisons
North America	Mean	0.005882	742
	SE	0.0002506	
	Median	0.003	
	SD	0.0068271	
	Minimum	0	
	Maximum	0.0345	
	Interquartile range	0.0108	
North America and Europe	Mean	0.010055	1665
	SE	0.000201	
	Median	0.0092	
	SD	0.0082036	
	Minimum	0	
	Maximum	0.0426	
	Interquartile range	0.0154	
Europe	Mean	0.005108	742
	SE	0.0002442	
	Median	0.0022	
	SD	0.0066511	
	Minimum	0	
	Maximum	0.0310	
	Interquartile range	0.0076	

containing genera with morphologically highly vs. poorly diagnosable species. These expected results suggest that DNA barcodes are effective identifiers regardless of species morphological characteristics and provide support for models of independent patterns of morphological and molecular evolution (Bromham *et al.* 2002; Davies & Savolainen 2006; Goldie *et al.* 2011).

DNA barcoding and biogeography

We found statistically significant differences when comparing intraspecific K2P genetic distances of individuals from a single continent, in our case from North America (NA) and Europe (E), to those calculated between NA and E (group NAE). The intercontinental median and mean values were 3/1.7 and 4.5/2 times greater compared with those from the groups NA and E, respectively, and single continental values overlapped greatly with those from NAE. The question arises whether DNA barcodes can still successfully identify spider species through expanded geographic scale over two discrete continents. Our results, counter to prediction, suggest yes. Applying the classic barcoding gap idea (Hebert *et al.* 2004b) to geographic scaling would predict the average K2P values to increase tenfold for the populations to be treated as separate species, and for barcoding to fail. In our case, the values increased, but much less than tenfold. Therefore, a North America and Europe genetic division does not preclude species identification using DNA barcoding.

The data set testing the effect of the geographic scale on barcoding consisted of different individuals from multiple families in addition to the three tested for the barcoding gap (Appendices 1 and 3). We therefore returned to our barcoding gap analysis to search for a possible overlap between the NAE intraspecific and those interspecific distances predicted by the barcoding gap analysis. Considering the highly family-specific average K2P genetic distances in that analysis, and a wider family coverage in the geographic data set, the logical comparison is not that of NAE with any specific family (e.g. NAE mean/median is 6/6.5 and 17.6/19.2 times lower than in the Lycosidae and Tetragnathidae, respectively) but rather with the 'all group' consisting of a three-family average data (Fig. 2). Such comparison found a slight overlap between the two groups, but the overlap disappeared at the 99.8th intraspecific (K2P = 0.031) and the 0.02th interspecific percentile (K2P = 0.036), suggesting a 99.8% species identification success over continents.

The geographic scale tested here was extreme, as the group NAE only included genetic distances between the two continents. Considering this, our test of the barcoding efficiency over continents was strict. On the other hand, the clear divide between the tested continents may provide a best-case example, and it may be possible that genetic distances would scale differently over more continuous geographic gradients, in particular in the tropics. Nevertheless, our testing suggests that barcodes almost always enable reliable species identification, yet they contain significant genetic variation over continents, which predicts DNA barcoding utility in biogeographic research.

Conclusions

Our results support models of independent patterns of morphological and molecular evolution by showing that DNA barcodes are effective in species identification regardless of their morphological diagnosibility. We conclude that despite containing informative biogeographic information, DNA barcodes represent an effective tool for identifying spider species over geographic (intercontinental) scales. However, the size of the barcoding gap strongly depends on taxonomic groups and practices. It is therefore necessary to establish a statistically better defined barcoding gap and its variation over taxonomic scales. Future studies should use proper statistical

© 2014 John Wiley & Sons Ltd

detection of the barcoding gap depending on actual data distribution.

Acknowledgements

This research was supported by the Slovenian Research Agency (Grants P1-0236 and MR-2013) and a Swiss Contribution to the enlarged EU grant (C1536-1 1T440013). For help or discussions, we thank Ingi Agnarsson, Andrej Blejec, Cene Fišer, Rok Kostanjšek, Tjaša Lokovšek, Ren-Chung Cheng, Matjaž Gregorič and Simona Kralj-Fišer. We also thank Vincent Nijman and two anonymous reviewers for helpful suggestions.

References

Agnarsson I, Coddington JA, Kuntner M (2013) Systematics: progress in the study of spider diversity and evolution. In: *Spider Research in the 21st Century: Trends and Perspectives* (ed. Penney D), pp. 58–111. Siri Scientific Press, Manchester, UK.

Aliabadian M, Kaboli M, Nijman V, Vences M (2009) Molecular identification of birds: performance of distance-based DNA barcoding in three genes to delimit parapatric species. *PLoS ONE*, **4**, e4119.

Arnedo MA, Ferrández MA (2007) Mitochondrial markers reveal deep population subdivision in the European protected spider *Macrothele calpeiana* (Walckenaer, 1805) (Araneae, Hexathelidae). *Conservation Genetics*, **8**, 1147–1162.

Ashfaq M, Hebert PDN, Mirza JH, Khan AM, Zafar Y, Mirza MS (2014) Analyzing mosquito (Diptera: Culicidae) diversity in Pakistan by DNA barcoding. *PLoS ONE*, **9**, e97268.

Avise JC (2000) *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge, Massachusetts, 447 pp.

Barrett RDH, Hebert PDN (2005) Identifying spiders through DNA barcodes. *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, **83**, 481–491.

Bell JR, Bohan DA, Shaw EM, Weyman GS (2005) Ballooning dispersal using silk: World fauna, phylogenies, genetics and models. *Bulletin of Entomological Research*, **95**, 69–114.

Bergsten J, Bilton DT, Fujisawa T *et al.* (2012) The effect of geographical scale of sampling on DNA barcoding. *Systematic Biology*, **61**, 851–869.

Bertolazzi P, Felici G, Weitschek E (2009) Learning to classify species with barcodes. *BMC Bioinformatics*, **10**, S7.

BioEssays (2009) DNA barcoding for ecologists. *Trends in Ecology & Evolution*, **24**, 110–117.

Blagoev G, Hebert P, Adamowicz S, Robinson E (2009) Prospects for using DNA barcoding to identify spiders in species-rich genera. *Zookeys*, **16**, 27–46.

Boykin LM, Armstrong KF, Kubatko L, Barro PD (2012) Species delimitation and global biosecurity. *Evolutionary Bioinformatics*, **8**, 1–37.

Bridge PD, Roberts PJ, Spooner BM, Panchal G (2003) On the unreliability of published DNA sequences. *New Phytologist*, **160**, 43–48.

Bromham L, Woolfit M, Lee MSY, Rambaut A (2002) Testing the relationship between morphological and molecular rates of change along phylogenies. *Evolution*, **56**, 1921–1930.

Burns JM, Janzen DH, Hajibabaei M, Hallwachs W, Hebert PDN (2007) DNA barcodes of closely related (but morphologically and ecologically distinct) species of skipper butterflies (Hesperiidae) can differ by only one to three nucleotides. *Journal of the Lepidopterists' Society*, **61**, 138–153.

Carr CM, Hardy SM, Brown TM, Macdonald TA, Hebert PDN (2011) A tri-oceanic perspective: DNA barcoding reveals geographic structure and cryptic diversity in Canadian polychaetes. *PLoS ONE*, **6**, e22232.

Collins RA, Boykin LM, Cruickshank RH, Armstrong KF (2012) Barcoding's next top model: an evaluation of nucleotide substitution models for specimen identification. *Methods in Ecology and Evolution*, **3**, 457–465.

Conn PB, McClintock BT, Cameron MF, Johnson DS, Moreland EE, Boveng PL (2013) Accommodating species identification errors in transect surveys. *Ecology*, **94**, 2607–2618.

Cummings MP, Neel MC, Shaw KL (2008) A genealogical approach to quantifying lineage divergence. *Evolution*, **62**, 2411–2422.

Davies TJ, Savolainen V (2006) Neutral theory, phylogenies, and the relationship between phenotypic change and evolutionary rates. *Evolution*, **60**, 476–483.

Delsinne T, Sonet G, Nagy ZT *et al.* (2012) High species turnover of the ant genus Solenopsis (Hymenoptera: Formicidae) along an altitudinal gradient in the Ecuadorian Andes, indicated by a combined DNA sequencing and morphological approach. *Invertebrate Systematics*, **26**, 457–469.

Dexter KG, Pennington TD, Cunningham CW (2010) Using DNA to assess errors in tropical tree identifications: how often are ecologists wrong and when does it matter? *Ecological Monographs*, **80**, 267–286.

Foelix R (2010) *Biology of Spiders*, 3rd edn. Oxford University Press, NY, New York.

Fujita MK, Leache AD, Burbrink FT, McGuire JA, Moritz C (2012) Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology and Evolution*, **27**, 480–488.

Goldie X, Lanfear R, Bromham L (2011) Diversification and the rate of molecular evolution: no evidence of a link in mammals. *Bmc Evolutionary Biology*, **11**, 286.

Haase P, Murray-Bligh J, Lohse S *et al.* (2006) Assessing the impact of errors in sorting and identifying macroinvertebrate samples. In: *The Ecological Status of European Rivers: Evaluation and Intercalibration of Assessment Methods* (eds Furse MT, Hering D, Brabec K, Buffagni A, Sandin L, Verdonschot PM). Developments in Hydrobiology 188, pp. 505–521. Springer, Dordrecht, the Netherlands.

Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN (2006) DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences, USA*, **103**, 968–971.

Hebert PDN, Barrett RDH (2005) Reply to the comment by L. Prendini on "Identifying spiders through DNA barcodes". *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, **83**, 505–506.

Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B-Biological Sciences*, **270**, 313–321.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004a) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences, USA*, **101**, 14812–14817.

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004b) Identification of birds through DNA barcodes. *Plos Biology*, **2**, 1657–1663.

Hebert PDN, deWaard JR, Landry J-F (2010) DNA barcodes for 1/1000 of the animal kingdom. *Biology Letters*, **6**, 359–362.

Hendrixson BE, DeRussy BM, Hamilton CA, Bond JE (2013) An exploration of species boundaries in turret-building tarantulas of the Mojave Desert (Araneae, Mygalomorphae, Theraphosidae, *Aphonopelma*). *Molecular Phylogenetics and Evolution*, **66**, 327–340.

Hull JM, Fish AM, Keane JJ, Mori SR, Sacks BJ, Hull AC (2010) Estimation of species identification error: Implications for raptor migration counts and trend estimation. *Journal of Wildlife Management*, **74**, 1326–1334.

Janecka J, Chowdhary B, Murphy W (2012) Exploring the correlations between sequence evolution rate and phenotypic divergence across the Mammalian tree provides insights into adaptive evolution. *Journal of Biosciences*, **37**, 897–909.

Johnsen A, Rindal E, Ericson PGP *et al.* (2010) DNA barcoding of Scandinavian birds reveals divergent lineages in trans-Atlantic species. *Journal of Ornithology*, **151**, 565–578.

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide-sequences. *Journal of Molecular Evolution*, **16**, 111–120.

Knowles LL, Carstens BC (2007) Delimiting species without monophyletic gene trees. *Systematic Biology*, **56**, 887–895.

© 2014 John Wiley & Sons Ltd

Kuntner M, Agnarsson I (2011a) Biogeography and diversification of hermit spiders on Indian Ocean islands (Nephilidae: *Nephilengys*). *Molecular Phylogenetics and Evolution*, **59**, 477–488.

Kuntner M, Agnarsson I (2011b) Phylogeography of a successful aerial disperser: the golden orb spider *Nephila* on Indian Ocean islands. *Bmc Evolutionary Biology*, **11**, 119.

Links MG, Dumonceaux TJ, Hemmingsen SM, Hill JE (2012) The chaperonin-60 universal target is a barcode for bacteria that enables de novo assembly of metagenomic sequence data. *PLoS ONE*, **7**, e49755.

Longhorn SJ, Nicholas M, Chuter J, Vogler AP (2007) The utility of molecular markers from non-lethal DNA samples of the cites II protected "tarantula" *Brachypelma vagans* (Araneae, Theraphosidae). *Journal of Arachnology*, **35**, 278–292.

Masters BC, Fan V, Ross HA (2011) Species delimitation–a Geneious plugin for the exploration of species boundaries. *Molecular Ecology Resources*, **11**, 154–157.

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

Meier R, Zhang G, Ali F (2008) The use of mean instead of smallest interspecific distances exaggerates the size of the "barcoding gap" and leads to misidentification. *Systematic Biology*, **57**, 809–813.

Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. *Plos Biology*, **3**, 1–10.

Negri M, Pileggi LG, Mantelatto FL (2012) Molecular barcode and morphological analyses reveal the taxonomic and biogeographical status of the striped-legged hermit crab species Clibanarius sclopetarius (Herbst, 1796) and Clibanarius vittatus (Bosc, 1802) (Decapoda: Diogenidae). *Invertebrate Systematics*, **26**, 561–571.

Nekola JC, White PS (1999) The distance decay of similarity in biogeography and ecology. *Journal of Biogeography*, **26**, 867–878.

Nijman V, Aliabadian M (2013) DNA barcoding as a tool for elucidating species delineation in wide-ranging species as Illustrated by owls (Tytonidae and Strigidae). *Zoological Science*, **30**, 1005–1009.

Nuñez JJ, Vejar-Pardo A, Guzmán BE, Barriga EH, Gallardo CS (2012) Phylogenetic and mixed Yule-coalescent analyses reveal cryptic lineages within two South American marine snails of the genus Crepipatella (Gastropoda: Calyptraeidae). *Invertebrate Biology*, **131**, 301–311.

Nwani CD, Becker S, Braid HE, Ude EF, Okogwu OI, Hanner R (2011) DNA barcoding discriminates freshwater fishes from southeastern Nigeria and provides river system-level phylogeographic resolution within some species. *Mitochondrial DNA*, **22**, 43–51.

Oliver I, Beattie AJ (1996) Invertebrate morphospecies as surrogates for species: a case study. *Conservation Biology*, **10**, 99–109.

O'Meara BC (2010) New heuristic methods for joint species delimitation and species tree inference. *Systematic Biology*, **59**, 59–73.

Platnick NI (2014) *The World Spider Catalog*. American Museum of Natural History, NY, New York.

Prendini L (2005) Comment on "Identifying spiders through DNA barcodes". *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, **83**, 498–504.

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

Rodrigo A, Bertels F, Heled J, Noder R, Shearman H, Tsai P (2008) The perils of plenty: what are we going to do with all these genes? *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*, **363**, 3893–3902.

Rosenberg NA (2007) Statistical tests for taxonomic distinctiveness from observations of monophyly. *Evolution*, **61**, 317–323.

Scott WA, Hallam CJ (2002) Assessing species misidentification rates through quality assurance of vegetation monitoring. *Plant Ecology*, **165**, 101–115.

Shea CP, Peterson JT, Wisniewski JM, Johnson NA (2011) Misidentification of freshwater mussel species (Bivalvia:Unionidae): contributing factors, management implications, and potential solutions. *Journal of the North American Benthological Society*, **30**, 446–458.

Smith FA (2008) Body size, energetics and evolution. In: *Encyclopedia of Ecology* (eds Jørgensen SE, Fath BD), pp. 477–482. Elsevier, Oxford.

Srivathsan A, Meier R (2012) On the inappropriate use of Kimura-2-parameter (K2P) divergences in the DNA-barcoding literature. *Cladistics*, **28**, 190–194.

Tamura K, Peterson D, Peterson N *et al.* (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Tavares ES, Goncalves P, Miyaki CY, Baker AJ (2011) DNA barcode detects high genetic structure within neotropical bird species. *PLoS ONE*, **6**, e28543.

Taylor HR, Harris WE (2012) An emergent science on the brink of irrelevance: a review of the past 8 years of DNA barcoding. *Molecular Ecology Resources*, **12**, 377–388.

Van Der Bank H, Herbert D, Greenfield R, Yessoufou K (2013) Revisiting species delimitation within the genus *Oxystele* using DNA barcoding approach. *Zookeys*, **365**, 337–354.

Vuataz L, Sartori M, Gattolliat JL, Monaghan MT (2012) Endemism and diversification in freshwater insects of Madagascar revealed by coalescent and phylogenetic analysis of museum and field collections. *Molecular Phylogenetics and Evolution*, **66**, 979–991.

Webster BL, Emery AM, Webster JP *et al.* (2012) Genetic diversity within Schistosoma haematobium: DNA barcoding reveals two distinct groups. *PLOS Neglected Tropical Diseases*, **6**, e1882.

Weitschek E, Velzen R, Felici G, Bertolazzi P (2013) BLOG 2.0: a software system for character-based species classification with DNA Barcode sequences. What it does, how to use it. *Molecular Ecology Resources*, **13**, 1043–1046.

White BP, Pilgrim EM, Boykin LM, Stein ED, Mazor RD (2014) Comparison of four species-delimitation methods applied to a DNA barcode data set of insect larvae for use in routine bioassessment. *Freshwater Science*, **33**, 338–348.

Wiemers M, Fiedler K (2007) Does the DNA barcoding gap exist? - A case study in blue butterflies (Lepidoptera: Lycaenidae). *Frontiers in Zoology*, **4**, 8.

Wright S (1943) Isolation by distance. *Genetics*, **28**, 114–138.

Yang Z, Rannala B (2010) Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences, USA*, **107**, 9264–9269.

Yassin A, Markow TA, Narechania A, O'Grady PM, DeSalle R (2010) The genus *Drosophila* as a model for testing tree- and character-based methods of species identification using DNA barcoding. *Molecular Phylogenetics and Evolution*, **57**, 509–517.

Zaldívar-Riverón A, Martínez JJ, Ceccarelli FS *et al.* (2010) DNA barcoding a highly diverse group of parasitoid wasps (Braconidae: Doryctinae) from a Mexican nature reserve. *Mitochondrial DNA*, **21**, 18–23.

Zhang C, Zhang DX, Zhu T, Yang Z (2011) Evaluation of a Bayesian coalescent method of species delimitation. *Systematic Biology*, **60**, 747–761.

Zuckerkandl E, Pauling L (1965) Molecules as documents of evolutionary history. *Journal of Theoretical Biology*, **8**, 357–366.

K.Č. and M.K. contributed equally to study design and writing. K.Č. performed data mining and analyses.

Data accessibility

All used data may be found in the online version of this article in Appendices S1–S3 and on Dryad doi: 10.5061/dryad.bv88g.

© 2014 John Wiley & Sons Ltd

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1 DNA barcodes used in the barcoding gap analysis.

Appendix S2 DNA barcodes used in the test of species identification over morphological scales.

Appendix S3 DNA barcodes used in the test of species identification over geographical scales.

© 2014 John Wiley & Sons Ltd