Molecular Ecology (2006) **15**, 3073–3082 doi: 10.1111/j.1365-294X.2006.02999.x

DNA barcoding reveals extraordinary cryptic diversity in an amphipod genus: implications for desert spring conservation

JONATHAN D. S. WITT,* DOUG L. THRELOFF† and PAUL D. N. HEBERT‡
Department of Biology, University of Waterloo, 200 University Ave. West, Waterloo, Ontario, Canada N2L 3G1, †883 Greensboro Road, Ventura, California, USA, ‡Department of Integrative Biology, University of Guelph, Guelph, Ontario, Canada N1G 2W1

Abstract

DNA barcoding has revealed unrecognized species in several animal groups. In this study we have employed DNA barcoding to examine *Hyalella*, a taxonomically difficult genus of amphipod crustaceans, from sites in the southern Great Basin of California and Nevada, USA. We assessed the extent of species diversity using a species screening threshold (SST) set at 10 times the average intrapopulation cytochrome *c* oxidase subunit I (COI) haplotype divergence. Despite the fact that this threshold approach is more conservative in delineating provisional species than the phylogenetic species concept, our analyses revealed extraordinary levels of cryptic diversity and endemism. The SST discriminated two provisional species within *Hyalella sandra*, and 33 provisional species within *Hyalella azteca*. COI nucleotide divergences among these provisional species ranged from 4.4% to 29.9%. These results have important implications for the conservation of life in desert springs — habitats that are threatened as a result of groundwater over-exploitation.

Keywords: 28S, conservation, cryptic species, cytochrome *c* oxidase I (COI), desert springs, *Hyalella*

Received 16 November 2005; revision accepted 21 April 2006

Introduction

The ability to identify species is fundamental to the development of effective conservation strategies. However, conservation programmes confront a serious challenge: in many groups of organisms, especially invertebrates, understanding of species boundaries is poor. This difficulty arises not only because the number of described species falls far short of actual diversity (Blaxter 2004), but also because species are often exceedingly difficult to recognize using morphological approaches. The prevalence of 'morphologically cryptic' species, diagnosable only with molecular studies (Knowlton 1993, 2000; Avise 2004), has added an additional problem that cannot be ignored. The biological literature is replete with examples of cryptic species, especially within invertebrate groups (e.g. Väinöla *et al*. 1994; King & Hanner 1998; Taylor *et al*. 1998; Müller 2000; Gómez *et al*. 2002; Liu *et al*. 2003; Pfenninger *et al*. 2003; Witt *et al*. 2003; Lee & Ó Foighil 2004; Penton *et al*. 2004; Stoks *et al*. 2005). Although the recognition of cryptic species is technically straightforward, they are seldom considered in biodiversity assessments owing to the lack of affordable diagnostic methods. Recent work, however, has suggested that DNA barcoding will provide conservation biologists with an inexpensive, simple tool to aid both identifications of known species and the recognition of undescribed ones.

A 648-bp segment of the mitochondrial cytochrome *c* oxidase I gene (COI) has been adopted as the DNA barcoding system for animal life (Hebert *et al*. 2003, 2004a, b). Savolainen *et al*. (2005) provide a good summary of DNA barcoding objectives, methodologies and rationale. As a result of declining costs and increasing ease of DNA sequencing, the development of a barcode library promises to create a system for the rapid identification of animal life. It will also provide a simple method to flag lineages that likely represent overlooked species. Past DNA barcoding studies have employed two approaches for recognizing overlooked taxa. One method involves the search for covariation between barcode divergences and life history or morphological characters. For example, COI barcoding in combination with information on larval morphology

Correspondence: Jonathan D. S. Witt, Fax: +1-519-746-0614; E-mail: jwitt@sciborg.uwaterloo.ca

and feeding preferences revealed 10 cryptic species in one lepidopteran genus (Hebert *et al*. 2004a). The second approach to species recognition involves the search for barcode divergences that are so large as to signal species differences with high probability. For example, Hebert *et al*. (2004b) suggested the presence of four overlooked species of North American birds, based on the discovery of deep barcode divergences within members of single putative species. They advocated the application of a standard screening threshold, set at 10 times the average intraspecific COI haplotype divergence among known species, to recognize new provisional species. The latter approach does require a well-established taxonomic system because if overlooked species constitute the majority, they inflate estimates of intraspecific variation. In the case of birds, this was not problematic as only 3% of species showed evidence of overlooked taxa. It must be emphasized that DNA barcoding does not seek to describe new species on the basis of sequence divergence patterns, but rather to flag lineages — provisional species that likely merit formal recognition.

In these prior studies, DNA barcoding has revealed unrecognized species, but under conditions that will not extend to all taxonomic groups. In particular, the information needed to assess covariation between COI and other characters may be unavailable. Also, by definition, the detailed taxonomic framework needed to determine average intraspecific divergences will not exist within groups where species diversity is poorly known. Does this mean that groups most in need of an alternative method of species recognition cannot be investigated? We argue that this is not the case, proposing that a threshold set at 10 times the average *intrapopulation* COI haplotype divergence can be a powerful tool for species recognition when the geographic scale of investigation is modest. We term this a species-screening threshold (SST) and argue that it can be effective in recognizing provisional species in little studied groups. Our decision to employ a 10 times SST (10× SST) reflects, in part, the prior adoption of a similar threshold value in birds (Hebert *et al*. 2004b). Aside from its obvious simplicity in application, the 10× SST is a conservative filter for species; it overlooks recently diverged taxa, while protecting against the artificial recognition of intraspecific variants as species. We justify employing *intrapopulation* variation as a surrogate for *intraspecific* variation in calculation of the threshold by restricting its application to populations from a small geographic region where the two values are likely to be congruent. Specifically, by focusing on local populations, the interpretational complications introduced by broad-scale phylogeographic patterns are avoided. Moreover, in the case of *Hyalella*, the effectiveness of this SST as a screening filter has empirical support, albeit in another geographic region. Prior allozyme studies on *Hyalella* populations in northern North America have provided a good sense of species boundaries and COI analysis on these populations reveal that intraspecific divergence values always fall below the 10× SST calculated for this region (Witt & Hebert 2000; Witt in preparation).

The test case

This study is focussed on the amphipod crustacean genus *Hyalella* that occurs in permanent freshwaters of North and South America. The genus has a complex taxonomic history, but currently includes three subgenera: *Austrohyalella* and *Mesohyalella* occur only in South America, while the subgenus *Hyalella* is restricted to North America (Bousfield 1996). The last subgenus contains nine described species; eight of these are very narrowly distributed — five are known from a single water body. By contrast, *Hyalella azteca sensu lato* is North America's most broadly distributed aquatic invertebrate, occurring from the Atlantic to Pacific coasts, and from Mexico to the arctic tundra. Allozyme and mitochondrial DNA (mtDNA) analyses have shown that *H. azteca* is a complex of at least seven cryptic species in eastern Canada (Hogg *et al*. 1998; Witt & Hebert 2000; Witt *et al*. 2003), and many additional undescribed species occur elsewhere (Witt *et al*. 2003; Wellborn *et al*. 2005; Witt, unpublished). These cryptic *H. 'azteca'* species occasionally occur in sympatry, providing evidence of their reproductive isolation. For example, in a study of joint allozyme and COI sequence variability, Witt & Hebert (2000) found up to three cryptic species in some locations. The possibility of such coexistence must be considered when estimating intrapopulation sequence divergences.

Hyalella populations in the southern Great Basin region of California and Nevada provide an interesting test case for the use of DNA barcoding in a regional appraisal of species diversity. The Great Basin was characterized by an anastomosing system of Pleistocene and Miocene pluvial lakes and rivers, which exist today as fragmented endoreic spring systems (Grayson 1993). These springs contain a highly endemic fauna that includes pupfishes, springsnails and hemipterans (Hershler & Sada 2002; Polhemus & Polhemus 2002; Smith *et al*. 2002; Liu *et al*. 2003), with many species confined to a single location. Although the geographically restricted distributions of these taxa have provoked serious conservation concerns, no ecological or natural historical observations exist for most hyalellid populations in this area.

In this study, we examine the utility of an SST equivalent to 10× the mean intrapopulation sequence divergence at COI for the recognition of provisional species within Great Basin hyalellids. We estimate and apply this threshold to populations of *H. 'azteca'*, as well as to populations of the narrow endemics *H. muerta* and *H. sandra*, both known from only two localities. A potential difficulty with this approach lies in determining what constitutes a population

at locations with two or more reproductively isolated lineages (Witt & Hebert 2000). As an operational criterion in this study, two (or more) clusters of haplotypes derived from individuals at a single site that exhibit more than 10% nucleotide sequence divergence from each other are treated as distinct populations. The validity of this operational criterion is subsequently tested by sequencing a subset of COI haplotypes for the nuclear 28S rDNA gene, or large ribosomal subunit (LSU). This gene is much more conserved than COI; closely related species often possess 28S sequences that are identical or nearly so (Taylor *et al*. 2002; Lee & Ó Foighil 2004). In fact, different amphipod genera regularly exhibit less than 2% sequence divergence at this gene (Cristescu & Hebert 2005). As a result, the 28S analysis will provide a highly conservative method of corroborating the ability of COI to discriminate reproductively isolated lineages that co-occur. In particular, the treatment of co-occurring lineages as distinct species will be validated if they are distinguished by 28S, and will be further reinforced if they are not sister taxa on the basis of this gene.

Methods

Specimens of *Hyalella azteca* were collected from 49 sites in the southern Great Basin in the Amargosa, Mojave, White, and Owens River drainages, as well as from locations in the Las Vegas Valley, Saline Valley, Steptoe Valley and the 'Area of Sterile Basins' (named by Hubbs & Miller 1948 for its impoverished fish fauna). Special emphasis was placed on springs in Ash Meadows, Nevada (Fig. 1, Table 1). Two endemic species, *Hyalella muerta* and *Hyalella sandra*, that are each known from only two springs in Death Valley National Park (Fig. 1, Table 1) were also collected. All specimens were stored in 95% ethanol from their time of capture until analysis.

Total DNA was extracted from four to 10 individuals from each population, as well as from a representative of the subgenus *Austrohyalella*, which was used as an outgroup. DNA was extracted by grinding a periopod in 50 μL of a proteinase K extraction buffer (Schwenk 1996). A 680-bp fragment of the mitochondrial COI gene was amplified using the primers LCO1490 and HCO2198 (Folmer *et al*. 1994). The 50-μL polymerase chain reactions (PCR) contained 0.5–1 μL (out of 50 μL) of DNA template, 5.0 μL 10× PCR buffer, 0.2 μM of each primer, 2.2 mM $MgCl_2$, 0.2 mM of each dNTP, and 1 unit of *Taq* DNA polymerase. The PCR conditions consisted of 1 min at 94 °C followed by 5 cycles of 60 s at 94 °C, 90 s at 45 °C, 60 s at 72 °C; followed by 35 cycles of 60 s at 94 °C, 90 s at 51 °C, 60 s at 72 °C; followed by 5 min at 72 °C. PCR products were gel purified (2% agarose) using the Qiaex kit (QIAGEN Inc.) and sequenced using the ABI PRISM BigDye terminator 3 sequencing kit (30 cycles, annealing at 55 °C). Products were sequenced in one direction using primer LCO1490. Electrophoresis was conducted on an ABI 377 automated sequencer (Applied Biosystems). The sequences were subsequently aligned using MEGA3 (Kumar *et al*. 2004). A subset of 33 individuals sequenced for COI were also sequenced for the 28S rDNA gene. An approximately 1400-bp fragment of 28S was amplified with the primer pair 28S-3311F GGGACTACCCCCTGAATTTAAGCAT, and 28S-4434R CCAGCTATCCTGAGGGAAACTTCG (see http://www.cladocera.uoguelph.ca/tools/). The 50-μL PCRs contained 1.0–2.0 μL (out of 50 μL) of DNA template, 5.0 μL 10× PCR buffer, 0.2 μM of each primer, 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, and 1 unit of *Taq* DNA polymerase. The PCR conditions consisted of 1 min at 94 °C followed by 39 cycles of 1 min at 94 °C, 1 min at 51 °C, 1 min at 72 °C; followed by 5 min at 72 °C. PCR products were again gel purified, and sequenced in both directions using the ABI PRISM BigDye terminator 3 sequencing kit (30 cycles, annealing at 50 °C); sequencing was conducted on an ABI 3730 automated sequencer (Applied Biosystems). The sequences were subsequently aligned using the CLUSTAL module in MEGA3 with a gap penalty of 10 and gap extension penalty of 3 (Hickson *et al*. 2000).

Neighbour-joining (NJ) analyses were conducted on the COI data set using Kimura's 2-parameter model (K2P) in MEGA3 to identify haplotypes. Unique COI sequences were then used to estimate the mean within-population pairwise sequence divergences (K2P) for all populations that possessed two or more haplotypes. The species-screening threshold (SST) was set at 10 times the average of these population estimates, and it was subsequently used to delineate provisional species.

To determine if COI and 28S provided congruent results with respect to lineages that co-occurred at the same site, separate maximum likelihood analyses were conducted on the 28S and COI data for 33 individuals that were analysed for both genes. The degree of congruence between COI and 28S was examined using the incongruence length difference test (Farris *et al*. 1994). This test was conducted in PAUP 4b10 (Swofford 2001) using 1000 heuristic search replicates with tree-bisection–reconnection (TBR) branch swapping, and the Multrees option invoked. The best-fit models selected by Akaike Information Criterion in MODELTEST 3.6 (Posada & Crandall 1998) were implemented in both cases, and both analyses were conducted using one heuristic search with the starting tree obtained by NJ and TBR branch swapping, as well as 10 additional replicates starting from a random tree. Nonparametric bootstrap support was estimated for both COI and 28S using 200 pseudo-replicates, with the starting tree for each obtained by NJ, and nearest neighbour interchange branch swapping in PAUP. Support was also estimated by conducting Bayesian analyses on both data sets using MRBAYES 3.0b4 (Huelsenbeck & Ronquist 2001). Each Bayesian analysis was run for 5 million generations sampling 50 000 trees using the model of sequence evolution estimated by MODELTEST, but with the parameters

Table 1 Sample sites in the Amargosa River drainage (A), area of Sterile Basins (ASB), Las Vegas Valley (LV), Mojave River drainage (M), Owens River drainage (OW), Steptoe Valley (S), Saline Valley (SV), and White River drainage (W). Locations in Ash Meadows are indicated with an asterisk*. The nominal species with sample sizes (*N*), as well as the provisional species and haplotypes that were detected within them (haplotypes) are given for each location. Haplotype designations follow those used in Fig. 2

Map locality	Sample site	Latitude/Longitude	Nominal species (*N*)	Provisional species (Haplotypes)
A1	South Indian Spring*	36°25′32″N/116°18′32″W	*Hyalella azteca* (5)	HaPS3 (12, 14)
A2	North Indian Spring*	36°25′33″N/116°18′32″W	*Hyalella azteca* (5)	HaPS3 (11–14)
A3	Marsh Spring*	36°28′46″N/116°19′33″W	*Hyalella azteca* (5)	HaPS3 (13–18)
A4	North Scruggs Spring*	36°25′58″N/116°18′31″W	*Hyalella azteca* (4)	HaPS3 (19–21)
A5	Point of Rocks Spring*	36°24′08″N/116°16′19″W	*Hyalella azteca* (5)	HaPS4 (22, 23)
A6	Devils Hole*	36°25′31″N/116°17′25″W	*Hyalella azteca* (6)	HaPS5 (24)
A7	Crystal Spring*	36°25′10″N/116°19′26″W	*Hyalella azteca* (5)	HaPS5 (25–27)
A8	Five Springs*	36°27′48″N/116°19′05″W	*Hyalella azteca* (4)	HaPS7 (30)
A9	Mary Scott Spring*	36°26′12″N/116°18′53″W	*Hyalella azteca* (5)	HaPS6 (28, 29)
A10	Rogers Spring*	36°28′46″N/116°19′31″W	*Hyalella azteca* (5)	HaPS8 (31, 32)
A11	Fairbanks Spring*	36°29′25″N/116°20′27″W	*Hyalella azteca* (5)	HaPS8 (33, 34)
A12	Warm Springs Canyon	35°58′04″N/116°55′50″W	*Hyalella azteca* (5)	HaPS9 (35–39)
A13	Grapevine Spring	37°01′13″N/117°22′55″W	*Hyalella azteca* (4)	HaPS10 (40–42)
A14	Surprise Spring	37°00′02″N/117°20′34″W	*Hyalella azteca* (7)	HaPS23 (91)
A15	Lower Vine Ranch Spring	37°01′06″N/117°23′12″W	*Hyalella azteca* (5)	HaPS23 (92–94)
A16	Travertine Springs	36°26′33″N/116°46′37″W	*Hyalella sandra* (5)	HsPS1 (76–79)
			Hyalella muerta (5)	–(1, 2, 4)
A17	Texas Springs	36°27′27″N/116°50′13″W	*Hyalella muerta* (5)	–(3)
A18	Nevares Springs	36°30′44″N/116°49′16″W	*Hyalella sandra* (5)	HsPS2 (80–83)
ASB	Side Hill Spring	38°15′21″N/116°41′19″W	*Hyalella azteca* (5)	HaPS25 (99, 100)
LV1	Lost Creek	36°09′22″N/115°29′43″W	*Hyalella azteca* (7)	HaPS33 (125–128)
LV2	Willow Spring	36°09′40″N/115°29′50″W	*Hyalella azteca* (6)	HaPS33 (121–124)
M1	Cedar Springs	34°18′32″N/117°18′57″W	*Hyalella azteca* (5)	HaPS13 (50, 51)
M2	Mojave River	34°30′50″N/117°15′41″W	*Hyalella azteca* (5)	HaPS13 (48, 51)
M3	Big Morongo Spring	34°02′53″N/116°34′01″W	*Hyalella azteca* (5)	HaPS13 (49)
M4	Afton Canyon Spring	35°02′09″N/116°23′09″W	*Hyalella azteca* (5)	HaPS13 (51)
OW1	Spring 94	37°11′38″N/118°11′57″W	*Hyalella azteca* (4)	HaPS22 (88, 89)
OW2	Spring 97	37°17′52″N/118°11′41″W	*Hyalella azteca* (5)	HaPS22 (84–87)
OW3	Tuttle Creek	36°34′19″N/118°06′29″W	*Hyalella azteca* (6)	HaPS31 (113–116)
OW4	Owens River	36°34′28″N/118°00′35″W	*Hyalella azteca* (5)	HaPS31 (110, 111)
OW5	Lubkin Creek	36°32′31″N/118°03′37″W	*Hyalella azteca* (6)	HaPS31 (110–112)
OW6	Spring 103	37°29′34″N/118°19′57″W	*Hyalella azteca* (5)	HaPS29 (108)
OW7	Big Spring	37°45′00″N/118°56′18″W	*Hyalella azteca* (5)	HaPS28 (105)
OW8	Fish Slough	37°31′01″N/118°23′59″W	*Hyalella azteca* (5)	HaPS24 (95–98)
OW9	Antelope Spring	37°19′52″N/118°05′22″W	*Hyalella azteca* (6)	HaPS22 (90)
OW10	Mule Spring	37°06′22″N/118°12′03″W	*Hyalella azteca* (5)	HaPS29 (106, 107)
OW11	BLM Spring	37°28′50″N/118°24′09″W	*Hyalella azteca* (5)	HaPS12 (44)
OW12	Warm Spring	37°16′00″N/118°16′15″W	*Hyalella azteca* (5)	HaPS30 (109)
S1	Steptoe Valley Spring	40°13′16″N/114°44′51″W	*Hyalella azteca* (9)	HaPS18 (67, 68), HaPS26 (101)
S2	Billy Pope Spring	39°37′40″N/114°54′18″W	*Hyalella azteca* (4)	HaPS20 (70–71)
S3	Grass Spring	39°31′07″N/114°54′50″W	*Hyalella azteca* (5)	HaPS21 (72–75)
SV	Saline Valley Marsh	36°41′37″N/117°49′42″W	*Hyalella azteca* (5)	HaPS13 (45–47)
W1	Moapa Warm Springs	36°42′05″N/114°42′58″W	*Hyalella azteca* (5)	HaPS14 (52–56)
W2	White River	37°27′38″N/115°11′36″W	*Hyalella azteca* (8)	HaPS19 (69)
W3	Blue Point Spring	36°23′21″N/114°25′56″W	*Hyalella azteca* (6)	HaPS11 (43)
W4	Hot Creek	38°22′38″N/115°08′51″W	*Hyalella azteca* (6)	HaPS15 (57–60)
W5	Comins Lake	39°09′54″N/114°48′42″W	*Hyalella azteca* (5)	HaPS16 (61)
W6	Hot Spring	38°58′01″N/115°04′35″W	*Hyalella azteca* (5)	HaPS17 (65, 66)
W7	Upper White River	38°54′36″N/115°03′35″W	*Hyalella azteca* (5)	HaPS1 (5–7)
W8	Spring 197	38°38′10″N/115°04′10″W	*Hyalella azteca* (10)	HaPS27 (102, 103)
				HaPS32 (117–120)
W9	Sunny Side Creek	38°25′16″N/115°04′28″W	*Hyalella azteca* (4)	HaPS27 (104)
W10	North Flagg Spring	38°25′26″N/115°01′15″W	*Hyalella azteca* (5)	HaPS2 (8–10)
W11	Meadow Valley Wash	37°39′11″N/114°30′02″W	*Hyalella azteca* (5)	HaPS17 (62–64)

Fig. 1 Map of study area showing collection sites and counties in California and Nevada, USA, within the Amargosa River drainage (A), area of Sterile Basins (ASB), Las Vegas Valley (LV), Mojave River drainage (M), Owens River drainage (OW), Saline Valley (SV), Steptoe Valley (S) and White River drainage (W).

estimated by MRBAYES. Each of five Markov chains (four cold and one hot) were run in both of these analyses, with the first 100 and 150 trees discarded as the burn-in as likelihood stationarity was achieved for COI and 28S within 10 000 and 15 000 generations, respectively. The partition frequencies among the remaining trees give the posterior probabilities, and provide an estimate of clade credibility.

Results

COI sequence diversity

Two hundred and eighty-two COI sequences were obtained from *Hyalella* individuals among the 52 locations (Fig. 1, Table 1). The 637-bp alignment contained 306 variable

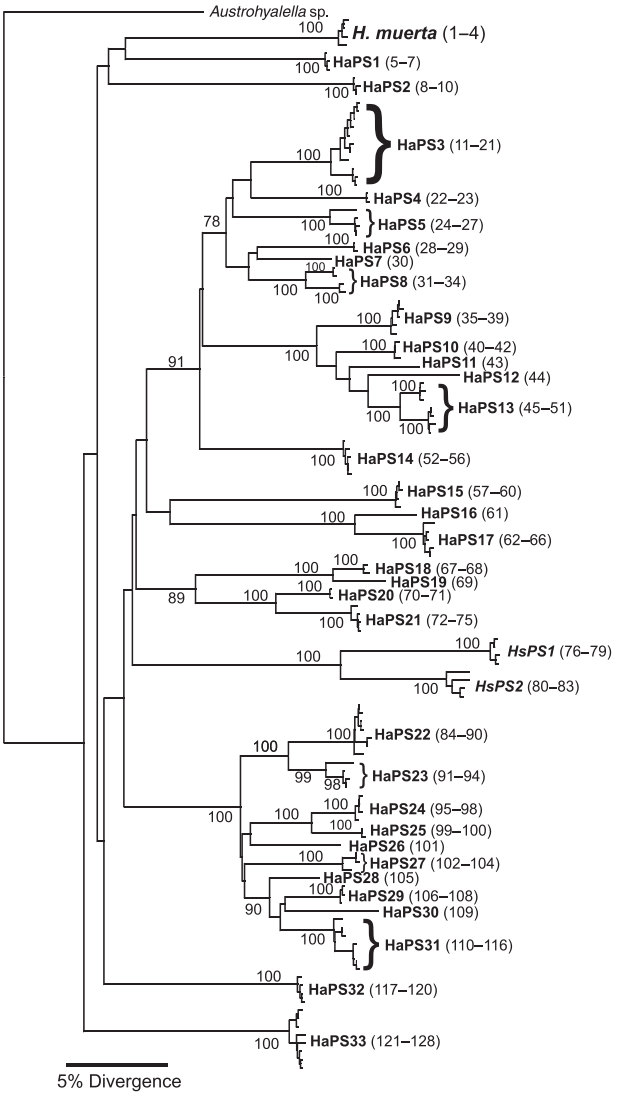

Fig. 2 Neighbour-joining phenogam (K2P) showing the relationships among 128 COI haplotypes derived from *Hyalella* populations in the Death Valley region. Bootstrap percentages (2000 pseudoreplicates) are given for nodes with greater than 70% support. Provisional species that are currently assigned to *Hyalella azteca* (HaPS) and *Hyalella sandra* (*HsPS*) identified by an SST of 3.75% nucleotide sequence divergence are indicated beside terminal branches. Numbers in parentheses identify haplotype designations within each provisional species, used elsewhere in the paper. For example, HaPS1 consists of three haplotypes, designated 5, 6 and 7 from top to bottom.

positions, 295 of which were parsimony informative, and amino acid sequence translations (invertebrate mitochondrial code) were unambiguous as there were no gaps or nonsense codons among the 128 haplotypes (GenBank accession numbers DQ464599–DQ464727, BOLD project: Great Basin Hyalellids, Process IDs GBH001-06 to GBH129-06) (Fig. 2). There was an extraordinary level of sequence variation; pairwise nucleotide sequence divergences among these haplotypes (K2P) were as high as 35.2%, while pairwise amino acid sequence divergences (Poisson correction) were as high as 18.2%.

Forty-seven of the 49 habitats contained a single lineage of *Hyalella azteca*, but two habitats, S1 and W8, contained two lineages that exhibited 22.8% and 23.7% nucleotide sequence divergence at COI, respectively (Table 1, Fig. 2). We treated these pairs of lineages as different populations in all subsequent analyses (see next section). Two or more COI haplotypes were detected in 39 populations (Table 1), and pairwise haplotype comparisons within these populations were used to calculate the mean intrapopulation sequence divergence. This value was 0.375% (n = 39, SE = 0.043%), producing an SST of 3.75%. The application of this SST indicated that the populations of *H. muerta* in Texas and Travertine Springs constitute a single species, but that the populations of *H. sandra* in Travertine and Nevares Springs represent two provisional species because they exhibit 15% sequence divergence. The 49 habitats with *H. azteca* are composed of 33 additional provisional species that exhibit pairwise COI nucleotide divergences ranging from a low of 4.4% (HaPS18 and HaPS19) to a high of 29.9% (Fig. 2). Populations in Ash Meadows were particularly diverse; six provisional species were detected within the 11 habitats in this spring complex (Fig. 2, Table 1). All provisional species clusters were well supported in the NJ analysis. In addition, a bootstrap analysis under the maximum parsimony criterion (100 pseudo-replicates with starting trees obtained by stepwise addition and nearest neighbour interchange branch swapping) with all characters weighted equally revealed similarly high levels of support. In this analysis, all provisional species haplotype clusters (clades) possess bootstrap support of 100%, with the exception of HaPS3, 8, 13, 18, and 23 that have bootstrap percentages of 99, 98, 84, 98 and 97, respectively. Although haplotypes within the provisional species invariably comprised a single well-supported monophyletic group, in some cases, such as HaPS13, they included two monophyletic clusters that exhibited less than 3.75% nucleotide sequence divergence (Fig. 2).

Phylogenetic analysis

The 28S sequences (GenBank accessions numbers DQ464728–DQ464760) ranged in length from 1382 to 1423 bp, and the sequence alignment was 1473 nucleotide positions in length. Nucleotide divergences among these 33 sequences (excluding *Austrohyalella*) were substantially lower than COI as expected, with a maximum of 5% (K2P). The incongruence length difference test did not provide statistical evidence for discordant signal between the COI and 28S data sets (P = 0.97). The best-fit models of nucleotide sequence substitution estimated by MODELTEST were the Tvm + I + G (α = 0.617, proportion of invariant sites = 0.470), which is a special case

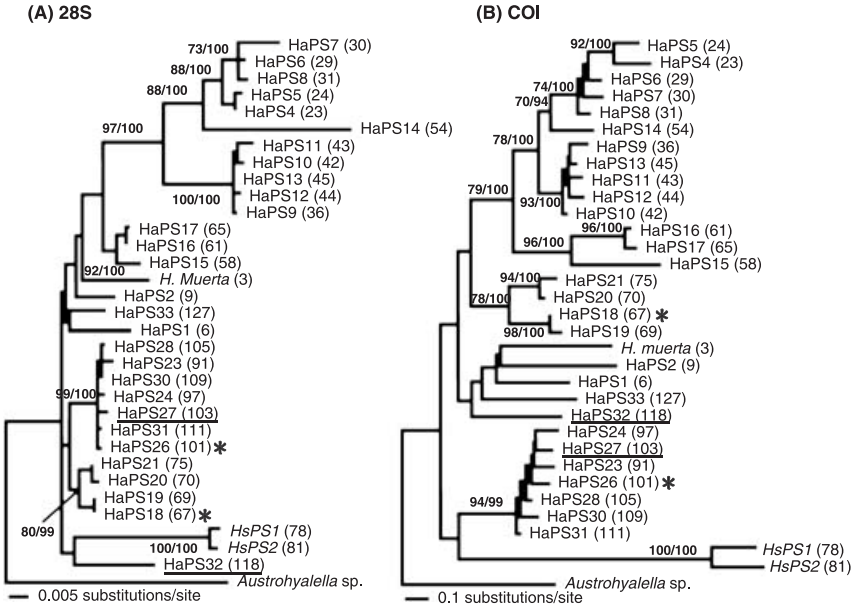

Fig. 3 Maximum likelihood phylograms for (A) 28S and (B) COI. Bootstrap support (200 pseudoreplicates) followed by posterior probabilities are given for nodes with greater than 70% bootstrap support. Provisional species are indicated at terminal branches with haplotype designations (see Fig. 2) in parentheses. Provisional species that were sympatric in S1 and W8 (Table 1) are respectively indicated with an asterisk (*) and underscored.

of the general time reversible model and GTR + I + G (α = 0.730, proportion of invariant sites = 0.610) for COI and 28S, respectively. Mean base frequencies among the COI and 28S sequences were, respectively, A: 0.24, C: 0.18, G: 0.23, T: 0.35 and A: 0.21, C: 0.26, G: 0.32, T: 0.21; there was no statistical evidence for heterogeneous nucleotide composition in either data set (χ^2 homogeneity test: $P = 0.09$, and 0.99). The phylograms estimated by maximum likelihood (Fig. 3) indicated general concordance between COI and 28S, and nodes that differ between them have low support in both analyses. For example, the 28S analysis suggested that the *H. 'sandra'* lineages are sister to HaPS32, while the COI analysis suggested an affinity to a different group of lineages (Fig. 3). Importantly, the 28S analysis corroborated the COI results for the lineages that co-occurred in S1 (HaPS18 and HaPS26) and W8 (HaPS27 and HaPS32) (Table 1). The 28S analysis discriminated the lineages within both sympatric pairs, and along with COI demonstrated that neither pair of co-occurring lineages represent sister taxa (Fig. 3).

Discussion

Our study has employed a species-screening threshold (SST) based on COI intrapopulation divergences as a surrogate measure for within-species variation. Species coexistence can create potential difficulties for the estimation of an SST because this would inflate apparent divergence values within populations. This was not problematic among Great Basin hyalellids as the only two cases of lineage co-occurrence involved taxa that were markedly diverged at COI (greater than 20% nucleotide sequence divergence), and discriminated by 28S. As a consequence, cases of species coexistence were readily recognized and treated as such. The within-population divergences resulted in an SST of 3.75%, a value slightly higher than the threshold divergence value (2.7%) used in an earlier study of birds (Hebert *et al*. 2004a). Despite this fact, the application of the SST revealed many overlooked species. *H. muerta* remained a single species, but the two known populations of *H. sandra* showed deep divergence, suggesting that they are actually different species. More strikingly, our results suggest that the Great Basin harbours at least 33 provisional species that are currently all assigned to *H. azteca*. The diverse hyalellid fauna revealed by our work appears to be largely endemic to the southern Great Basin as extensive surveys across North America, involving the sequencing of over 2000 *Hyalella* specimens (Witt & Hebert 2000; Witt *et al*. 2003; Witt, in progress), have revealed that only three of these taxa occur outside the study area — HaPS25 is widespread in North America, while HaPS13 has been detected on Catalina Island and in coastal California rivers and HaPS17 is known from Oregon and Washington.

Objections to the use of an SST to define provisional species may arise because it does not incorporate an explicit species concept. However, DNA barcoding seeks to provide a method to easily recognize described species and to *provisionally* recognize undescribed ones, not to define them. Coyne & Orr (2004) note that the criteria used to define species need not be the same as those used to recognize them. Moreover, it is important to emphasis that SSTs and DNA barcodes do not themselves constitute a species concept as erroneously suggested by some authors. Rather, they are tools to discriminate biological units, provisional species, which likely satisfy standard species criteria and should be subject to additional investigation. In fact, the

3.75% threshold used in this study provided results consistent with a phylogenetic species concept since the provisional species recognized through its application were all well-supported monophyletic groups. Strict phylogenetic species criteria (e.g. Cracraft 1983, 1989) have been criticized for over-splitting taxa to avoid paraphyletic species (Harrison 1998; Avise 2004), resulting in inconsistencies with biological and morphological species criteria (see Templeton 1998, discussion of 'speciation by remote control'). We note that the SST approach adds an important element of conservatism to the recognition of provisional species that is lacking in the implementation of a purely phylogenetic species criterion. Hence, some provisional species identified here (e.g. HaPS8 and HaPS13) include well-supported haplotype clusters that would be recognized as different species under the phylogenetic concept, but not here because their divergences are below the SST. In short, not all provisional species consisted of the 'smallest diagnosable cluster' stipulated by the phylogenetic species concept. In fact, the lowest magnitude of nucleotide sequence divergence between any provisional species pair recognized in this study was 4.4% (HaPS18 and HaPS19). The 3.75% threshold that we employed is consistent with insights gained through the study of reproductive isolation in *Hyalella* populations in other regions of North America. Combined allozyme and COI surveys in over 70 *Hyalella* populations from northern North America (Witt & Hebert 2000; Witt, in progress) revealed that fixed allozyme differences, which provide strong evidence for reproductive isolation, ordinarily only occur among sympatric lineages exhibiting 3% or more sequence divergence at COI. We recognize that our SST threshold may have overlooked recently diverged biological species among Great Basin hyalellids. However, if this has been the case, traditional taxonomic approaches have clearly performed much worse.

The present results have important implications for aquatic conservation in the deserts of the southern Great Basin. Aquatic habitats in this region are both uncommon and small. Because water is in short supply for agricultural and domestic use (Sada & Vinyard 2002), many springs in this area have disappeared altogether as a result of groundwater exploitation (Shepard 1993; Nash 2005). Sada & Vinyard (2002) examined the population status of 199 endemic aquatic organisms in the Great Basin, including 104 vertebrates. They found that 50% of these taxa have suffered population losses and 16 have become extinct, all as a consequence of groundwater exploitation. Importantly, they also show that habitat size is an important indicator of both endemism and vulnerability; small springs harbour numerous endemics and they are also most at risk. Some springs have received special conservation status as a result of endemic vertebrates (e.g. the pupfish *Cyprinodon diabolis* in Devils Hole). Ironically, efforts to conserve such fishes have sometimes involved their translocation to new locations with unknown impacts on invertebrate faunas that may have been rich in endemics. The present study, as well as that of Liu *et al*. (2003), strongly suggests that diversity and endemism in the invertebrate faunas of Great Basin spring systems is far higher than previously realized. Clearly this 'cryptic fauna' merits consideration in conservation efforts. Surprisingly, the ability of DNA barcoding to discriminate unrecognized diversity has caused some authors concern. Ebach & Holdredge (2005) lament that 'barcoding will almost certainly result in a plethora of newly "flagged" DNA species that will never formally be described', and suggest that it may take 250 years for traditional taxonomy to catch up with barcoding. Meyer & Paulay (2005) offer a similarly pessimistic view of the use of sequence divergence thresholds in taxonomically understudied groups. Despite their 83% success rate in species delineation among marine cowries, they suggest that thresholds 'do not bode well for delineating closely related species in taxonomically under studied groups', and further suggest that the successful discrimination of species in groups such as *Hyalella* will require a centuries long initial phase of traditional taxonomic consideration. These arguments fail to recognize that threats to biodiversity in the Great Basin (Grayson 1993; Shepard 1993; Sada & Vinyard 2002; Nash 2005) and beyond are real and immediate. Because DNA barcoding can provide a rapid assessment of species diversity rather than a millennial wait, it is well positioned to aid the development of strategies that minimize biodiversity loss.

Acknowledgements

We thank Angela Hollis for assisting with sequencing, and the staff at Death Valley National Park for field assistance. Funding for this study was provided by grants from NSERC and the Canada Research Chairs program to P.D.N.H. and an NSERC Postdoctoral Fellowship to J.W.

References

Avise JC (2004) *Molecular Markers Natural History and Evolution*. Sinauer Associates, Sunderland, Massachusetts.

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

Bousfield EL (1996) A contribution to the reclassification of neotropical freshwater hyalellid amphipods (Crustacea: Gammaridea, Talitroidea). *Bolletino del Museo Civico di Storia Naturale di Verona*, **20**, 175–224.

Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Sunderland, Massachusetts.

Cracraft J (1983) Species concepts and speciation analysis. In: *Current Ornithology* (ed. Johnstone RF), pp. 159–187. Plenum, New York.

Cracraft J (1989) Speciation and its ontology: the empirical consequences of alternative species concepts for understanding pattern and process of differentiation. In: *Speciation and its*

Consequences (eds Otte D, Endler JA), pp. 28–59. Sinauer Associates, Sunderland, Massachusetts.

Cristescu ME, Hebert PDN (2005) The 'Crustacean Seas' — an evolutionary perspective on the Ponto-Caspian peracarids. *Canadian Journal of Fisheries and Aquatic Sciences*, **62**, 505–512.

Ebach MC, Holdredge C (2005) DNA barcoding is no substitute for taxonomy. *Nature*, **434**, 697.

Farris JS, Kallersjo M, Kluge AG, Bult C (1994) Testing significance of incongruence. *Cladistics*, **10**, 315–319.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, **3**, 294–299.

Gómez A, Serra M, Carvalho G, Lunt H (2002) Speciation in ancient cryptic species complexes: evidence from the molecular phylogeny of *Brachionus plicatilis* (Rotifera). *Evolution*, **56**, 1431–1444.

Grayson DK (1993) *The Deserts's Past: A Natural Prehistory of the Great Basin*. Smithsonian Institution Press, Washington, DC.

Harrison RH (1998) Linking evolutionary patterns and process. The relevance of species concepts for the study of speciation. In: *Endless Forms, Species and Speciation* (eds Howard DJ, Berlocher, SH), pp. 19–31. Oxford University Press, New York.

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

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

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004b) 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.

Hershler R, Sada DW (2002) Biogeography of Great Basin aquatic snails of the Genus *Pyrgulopsis*. In: *Great Basin Aquatic Systems History* (eds Hershler R, Madsen DB, Currey DR), pp. 255–276. Smithsonian Contributions to the Earth Sciences, 33. Smithsonian Institution Press, Washington DC.

Hickson RE, Simon C, Perrey SW (2000) The performance of several multiple sequence alignment programs in relation to secondary structure features for an rRNA sequence. *Molecular Biology and Evolution*, **13**, 3527–3542.

Hogg ID, Celine L, de Lafontaine Y, Doe KG (1998) Genetic evidence for a *Hyalella* species complex within the Great Lakes-St. Lawrence River drainage basin: implications for ecotoxicology and conservation biology. *Canadian Journal of Zoology*, **76**, 1134–1140.

Hubbs CL, Miller RR (1948) The zoological evidence: correlation between fish distribution and hydrographic history in the desert basins of the western United States. In the Great Basin, with emphasis on glacial and postglacial times. *Bulletin of the University of Utah*, **38**, 17–166.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, **17**, 745–755.

King JK, Hanner R (1998) Cryptic species in a 'living fossil' lineage: taxonomic and phylogenetic relationships within the genus *Lepidurus* (Crustacea: Notostraca) in North America. *Molecular Phylogenetics and Evolution*, **13**, 23–36.

Knowlton N (1993) Sibling species in the sea. *Annual Review of Ecology and Systematics*, **24**, 189–216.

Knowlton N (2000) Molecular genetic analyses of species boundaries in the sea. *Hydrobiologia*, **420**, 73–90.

Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Bioinformatics*, **5**, 150–163.

Lee T, Ó Foighil D (2004) Hidden Floridian biodiversity: mitochondrial and nuclear gene trees reveal cryptic species within the scorched mussel, *Brachidontes exustus*, species complex. *Molecular Ecology*, **13**, 3525–3542.

Liu H-P, Hershler R, Clift K (2003) Mitochondrial DNA sequences reveal extensive cryptic diversity within a western American spring snail. *Molecular Ecology*, **12**, 2771–2782.

Meyer CP, Paulay G (2005) DNA Barcoding: error rates based on comprehensive sampling. *PLOS Biology*, **3**, 2229–2238.

Müller J (2000) Mitochondrial DNA variation and the evolutionary history of cryptic *Gammarus fossarum* types. *Molecular Phylogenetics and Evolution*, **15**, 260–268.

Nash JM (2005) Western water wars. *Time Magazine*, **166**, 18–19.

Penton EH, Hebert PDN, Crease TJ (2004) Mitochondrial DNA variation in North American populations of *Daphnia obtusa*: continentalism or cryptic endemism. *Molecular Ecology*, **13**, 97–107.

Pfenninger M, Staubach S, Albrecht C, Schwenk K (2003) Ecological and morphological differentiation among cryptic evolutionary lineages in freshwater limpets of the nominal form-group *Ancylus fluviatilis* (O.F. Müller, 1774). *Molecular Ecology*, **12**, 2731–2745.

Polhemus DA, Polhemus JT (2002) Basins and Ranges: The biogeography of aquatic true bugs (Insecta; Heteroptera) in the Great Basin. In: *Great Basin Aquatic Systems History* (eds Hershler R, Madsen DB, Currey DR), pp. 235–254. Smithsonian Contributions to the Earth Sciences, 33.

Posada D, Crandall KA (1998) Testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.

Sada DW, Vinyard GL (2002) Anthropogenic changes in biogeography of Great Basin aquatic biota. In: *Great Basin Aquatic Systems History* (eds Hershler R, Madsen DB, Currey DR), pp. 75–234. Smithsonian Contributions to the Earth Sciences, 33.

Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R (2005) Toward writing the encyclopaedia of life: an introduction to DNA barcoding. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **360**, 1805–1811.

Schwenk K (1996) *Evolutionary Genetics of Daphnia Species Complex: Hybridism in Syntopy*. Publication 2231. Netherlands Institute of Ecology, Centre for Limnology, Neuwersluis, The Netherlands.

Shepard WD (1993) Desert springs-both rare and endangered. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **3**, 351–359.

Smith GR, Dowling TE, Gobalet KW, Lugaski T, Shiozawa DK, Evans RP (2002) Biogeography and timing of evolutionary events among Great Basin fishes. In: *Great Basin Aquatic Systems History* (eds Hershler R, Madsen DB, Currey DR), pp. 175–234. Smithsonian Contributions to the Earth Sciences, 33.

Stoks R, Nystrom JL, May ML, McPeek MA (2005) Parallel evolution in ecological and reproductive traits to produce cryptic damselfly species across the Holarctic. *Evolution*, **59**, 1976–1988.

Swofford DL (2001) PAUP*: *Phylogenetic analysis using parsimony, Version 4.0b10*. Sinauer Associates, Sunderland, Massachusetts.

Taylor DJ, Finston TL, Hebert PDN (1998) Biogeography of a widespread freshwater crustacean: pseudocongruence and cryptic endemism in the North American *Daphnia laevis* complex. *Evolution*, **52**, 1648–1670.

Taylor DJ, Ishikane CR, Haney RA (2002) The systematics of the Holarctic bosminids and a revision that reconciles molecular and morphological evolution. *Limnology and Oceanography*, **47**, 1486–1495.

Templeton AR (1998) Species and speciation: geography, population structure, ecology and gene trees. In: *Endless Forms, Species and Speciation* (eds Howard DJ, Berlocher SH), pp. 32–43. Oxford University Press, New York.

Väinöla R, Riddoch BJ, Ward RD, Jones RI (1994) Genetic zoogeography of the *Mysis relicta* species group (Crustacea: Mysidacea) in northern Europe and North America. *Canadian Journal of Fisheries and Aquatic Sciences*, **51**, 1490–1505.

Wellborn GA, Cothran R, Bartholf S (2005) Life history and allozyme diversification in regional ecomorphs of the *Hyalella azteca* (Crustacea: Amphipoda) species complex. *Biological Journal of the Linnean Society*, **84**, 161–175.

Witt JDS, Hebert PDN (2000) Cryptic species diversity and evolution in the amphipod genus *Hyalella* within central glaciated North America: a molecular phylogenetic approach. *Canadian Journal of Fisheries and Aquatic Sciences*, **57**, 687–698.

Witt JDS, Blinn DW, Hebert PDN (2003) The recent evolutionary origin of the phenotypically novel amphipod *Hyalella montezuma* offers an ecological explanation for morphological stasis in a closely allied species complex. *Molecular Ecology*, **12**, 405–413.

Jonathan Witt's research interests include rates of morphological and molecular divergence, biogeography and biodiversity of aquatic organisms. Doug Threloff's interests are focused on the natural history and biodiversity of desert ecosystems. Paul Hebert's research is focused on DNA-based identification systems and global biodiversity.
