

RESEARCH ARTICLE

Open Access

Mitochondrial and nuclear phylogenetic analysis with Sanger and next-generation sequencing shows that, in Área de Conservación Guanacaste, northwestern Costa Rica, the skipper butterfly named *Urbanus belli* (family HesperIIDae) comprises three morphologically cryptic species

Claudia Bertrand¹, Daniel H Janzen², Winnie Hallwachs², John M Burns³, Joel F Gibson¹, Shadi Shokralla^{1,4} and Mehrdad Hajibabaei^{1*}

Abstract

Background: Skipper butterflies (HesperIIDae) are a relatively well-studied family of Lepidoptera. However, a combination of DNA barcodes, morphology, and natural history data has revealed several cryptic species complexes within them. Here, we investigate three DNA barcode lineages of what has been identified as *Urbanus belli* (HesperIIDae, Eudaminae) in Área de Conservación Guanacaste (ACG), northwestern Costa Rica.

Results: Although no morphological traits appear to distinguish among the three, congruent nuclear and mitochondrial lineage patterns show that "*Urbanus belli*" in ACG is a complex of three sympatric species. A single strain of *Wolbachia* present in two of the three cryptic species indicates that *Urbanus segnestami* Burns (formerly *Urbanus bellidHJ01*), *Urbanus bernikemi* Burns (formerly *Urbanus bellidHJ02*), and *Urbanus ehakernae* Burns (formerly *Urbanus bellidHJ03*) may be biologically separated by *Wolbachia*, as well as by their genetics. Use of parallel sequencing through 454-pyrosequencing improved the utility of *ITS2* as a phylogenetic marker and permitted examination of the intra- and interlineage relationships of *ITS2* variants within the species complex. Interlineage, intral lineage and intragenomic compensatory base pair changes were discovered in the secondary structure of *ITS2*.

Conclusion: These findings corroborate the existence of three cryptic species. Our confirmation of a novel cryptic species complex, initially suggested by DNA barcode lineages, argues for using a multi-marker approach coupled with next-generation sequencing for exploration of other suspected species complexes.

Keywords: Lepidoptera, Intragenomic variation, Non-metric multi-dimensional scaling, Phylogeny, DNA barcoding

* Correspondence: mhajibab@uoguelph.ca

¹Biodiversity Institute of Ontario & Department of Integrative Biology, University of Guelph, Guelph, Ontario N1G 2W1, Canada
Full list of author information is available at the end of the article

Background

The identification of species is essential for robust biodiversity estimates and a wide range of biological investigations with socio-economic applications. For example, the inability to accurately detect species has historically led to costly invasive species outbreaks and sub-optimal conservation practices for species at risk [1-3]. Herein we focus on “cryptic” species and define them according to Bickford’s overarching definition: two or more distinct species that are erroneously classified under one species name [1]. Morphologically cryptic species—those species that lack anatomical differences—require additional lines of evidence for species identification beyond traditional measures, including but not limited to: ecological distinctiveness, behavioral differences, morphological dissimilarity in undescribed life stages, and molecular divergences e.g., [4-6]. This study focuses on the latter and, more specifically, utilizes DNA barcode divergences as initial proxies for distinguishing cryptic species. DNA barcoding utilizes a short, standardized gene region to identify species; this study uses the 658 bp region of the cytochrome *c* oxidase I gene (*COI*), chosen as a DNA barcode for animal life [7]. Since 2003, DNA barcoding has been applied to a long-standing morphological and ecological inventory of Lepidoptera from the 140,000 ha terrestrial portion of Área de Conservación Guanacaste (ACG), northwestern Costa Rica (<http://www.acguanacaste.ac.cr>).

The ACG Lepidoptera inventory is comprehensively collecting and cataloging ancillary data including: DNA barcodes, food plants, parasitoids, fine-scale geographic and ecosystem data, and adult and caterpillar morphology [8] (<http://janzen.sas.upenn.edu>). Numerous instances of otherwise unrecognized species diversity have been discovered within the ACG. Several studies have focused on cryptic species complexes within the skipper butterflies (Hesperiidae) [9-13].

The discovery and characterization of species of hesperiid butterflies in ACG includes one of the most highly cited studies using DNA barcoding [13]. This study involved the *Astraptes fulgerator* (Walch, 1775) species complex and found ten distinct DNA barcode lineages. Although, some of these lineages displayed low sequence divergence (e.g., less than 1%), they were confirmed as species through the strong correlation of barcodes with distinct food plants, caterpillar coloration, some subtle differences in adult coloration and size, and ecological preferences [13].

In cases where cryptic species have been detected in Lepidoptera, natural history traits that correlate with *COI* lineages are key diagnostic characters. In this study, we explore a novel case where natural history traits appear to be absent, yet marked (i.e., more than 2%) DNA barcode divergences are present. We focus our efforts on what has been called “*Urbanus belli*” [11] in the ACG

and explore the utility of a genetic framework to identify morphologically cryptic species.

“*Urbanus belli*” is viewed as well-categorized taxonomically and ecologically as a single species that has become specialized to feed on the leaves of tropical herbaceous or shrubby Asteraceae, i.e., composites [11] (<http://janzen.sas.upenn.edu>). Within the small area of ACG, this skipper butterfly contains three distinct DNA barcode lineages that have temporarily been called *Urbanus belli*DHJ01, *Urbanus belli*DHJ02, and *Urbanus belli*DHJ03 in [11]. They are separated from each other by ~3-5% *COI* sequence divergence (Figure 1). As mentioned previously, these three lineages lack apparent diagnostic morphological and natural history characters, other than that one of them, within ACG, is restricted to rain forest, where all three lineages occur naturally (Figure 2). Food plant lists overlap among all three lineages (Table 1).

To test the hypothesis that “*Urbanus belli*” in the ACG is a complex of three cryptic species, as suggested by *COI* data (Figure 1), we compared lineage patterns across three additional molecular markers including: another mitochondrial gene (*cytochrome b*; *cytb*), a protein-coding nuclear gene (*elongation factor subunit 1 alpha*; *EF1 α*), and a nuclear ribosomal region (*Internal transcribed spacer region II*; *ITS2*), through the construction of phylogenetic trees. Given the high copy number and varying levels of intragenomic heterogeneity of the *ITS2* [14,15], Next-Generation Sequencing (NGS) was used to obtain sequences of *ITS2* from each specimen and non-metric multidimensional scaling (nMDS) was used to analyze intragenomic, intraspecific, and interspecific sequence clustering.

Additional analyses, including compensatory base changes (CBCs) in *ITS2* secondary structure and a *Wolbachia* endosymbiosis assay, were performed to supplement results in light of the absence of natural history characters. CBCs in the secondary structure of *ITS2* have been found to correlate strongly with distinct biological species of plants and fungi and as such, have been proposed as a molecular indicator of biological species [15-17]. The occurrence of a single CBC has been shown to correlate with distinct species of plants 93.11% of the time [16,17]. Lastly, the presence of *Wolbachia* endosymbionts was tested by using the *Wolbachia surface protein* (*wsp*) marker followed by Multi Locus Sequence Typing (MLST) [18]. *Wolbachia* are known to impact insect evolution by altering host reproduction. The most commonly reported mechanism is cytoplasmic incompatibility [19], the inability of infected males to reproduce successfully with uninfected females. Cytoplasmic incompatibility has been a factor in the reduction of gene flow between infected and non-infected populations, populations infected with different strains of *Wolbachia*, and populations infected with the same strain of *Wolbachia* [20]. By including this analysis, we begin to

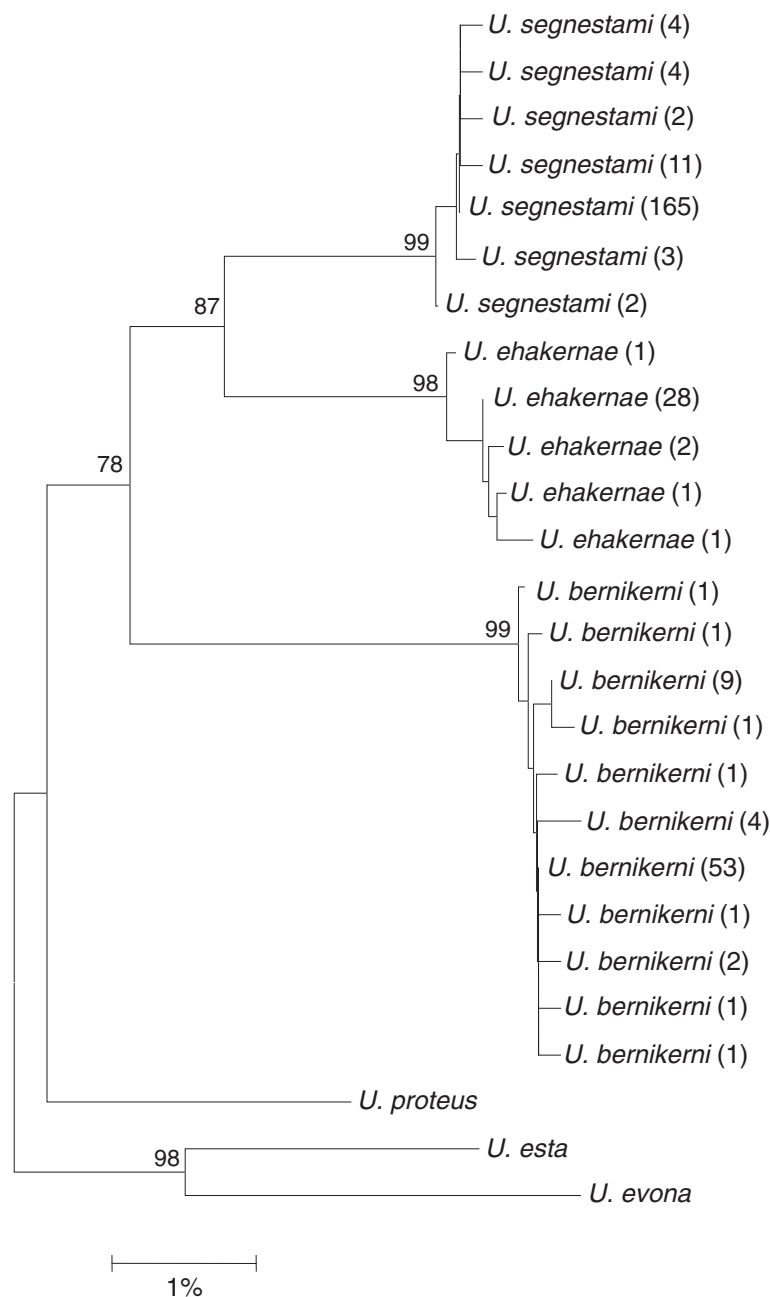


Figure 1 NJ-tree of 299 *Urbanus bernikerni* complex sequences from BOLD (<http://boldsystems.org>). Sequence divergence was estimated using the Kimura two-parameter method. Nodal support is based on 1000 bootstrap replicates. Scale bar represents number of substitutions per site. Numbers in parentheses are the number of individuals with each kind of haplotype.

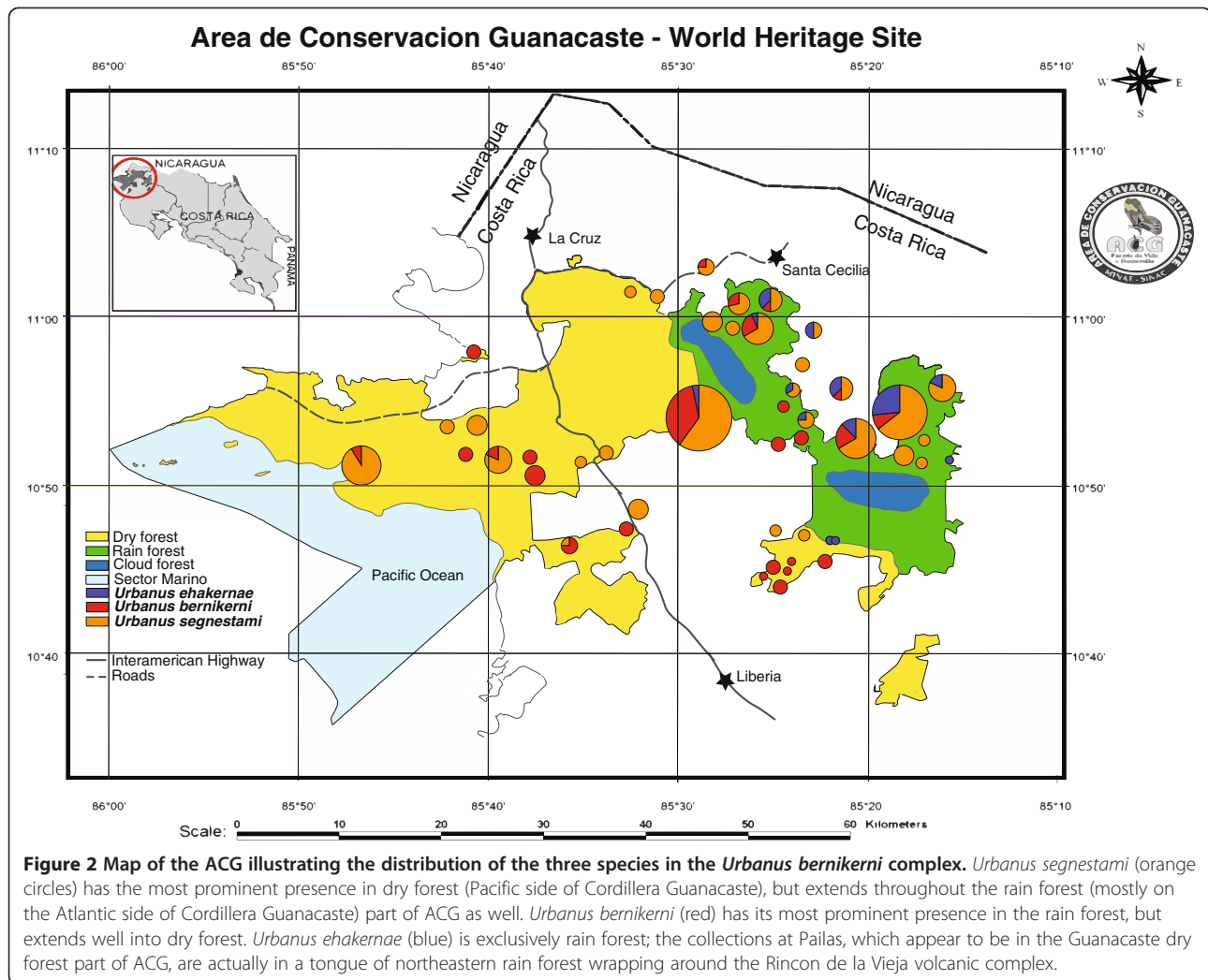
explore the role that *Wolbachia* infections can play as potential reinforcing mechanisms for species' isolation [20].

Methods

Specimens

Wild caterpillars were collected from ACG, reared to adults, and added to the ongoing, comprehensive inventory of Lepidoptera, with ancillary data, including locality, food

plant species, sex, and parasitoids [8,21]. Eight hundred and fifty wild-caught "*Urbanus belli*" caterpillars have been reared by the ACG inventory, and 325 of the resulting adults have now been barcoded. Table 1 provides ancillary data for all 325 adults. Fifty-one DNA barcoded individuals were selected from the Barcode of Life Data Systems (BOLD) [22] for investigation in this study. All specimens have been morphologically identified as "*Urbanus belli*."



Molecular analysis

DNA was extracted from a single leg of 51 specimens, using a Nucleospin Tissue Kit (Macherey-Nagel Inc., Bethlehem, PA, USA). PCR amplification and sequencing was performed for four gene regions: *COI*, *cytb*, *EF1 α* , and *ITS2* (Table 2, amplification primers). *EF1 α* was amplified using four primer sets that overlapped by 20-30 bp (Table 2). *COI*, *cytb*, and each fragment of *EF1 α* were amplified in 25 μ L reactions containing final concentrations of 0.2 mM of dNTPs, 0.2 μ M of each primer, 0.024 U/ μ L *Taq* DNA polymerase (Invitrogen, Life Technologies, Burlington, ON, Canada), 20 μ M of Tris-HCl, 50 μ M KCl, 2.5 mM MgCl₂, and 30-100 ng of DNA template. The PCR thermal cycling conditions varied by marker with *COI* and *EF1 α* : 94°C for 2 min., 40 cycles (94°C for 30 sec., 51°C for 40 sec., 72°C for 1 min.), 72°C for 5 min. Conditions for *Cytb* were: 94°C for 2 min., 35 cycles (94°C for 30 sec., 48°C for 1 min., 72°C for 40 sec.), 72°C for 1 min (primer information see Table 2).

ITS2 was amplified in 25 μ L reactions containing final concentrations of 0.2 mM of dNTPs, 0.4 μ M of each primer, 0.024 U/ μ L *Taq* DNA polymerase, 20 μ M of Tris-HCl, 50 μ M KCl, 1.5 mM MgCl₂, and 30-100 ng of DNA template. The PCR thermal cycling conditions were 94°C for 5 min., 40 cycles (94°C for 1 min., 53°C for 1 min., 72°C for 1 min.), 72°C for 10 min.

The PCR results were visualized on a 1.5% agarose gel stained with ethidium bromide. Any PCR products that showed a band on the agarose gel underwent cycle sequencing reactions with the following protocol: 0.25 μ L of Dye terminator mix v3.1, 1.875 μ L of 5 \times Sequencing Buffer, 1 μ L of 10 μ M Primer, 5.875 μ L of H₂O, and 2.0 μ L of PCR product for a final volume of 11 μ L. The cycle sequencing thermocycling protocol was as follows: 96°C for 2 min., 30 cycles (96°C for 30 sec., 55°C for 15 sec., 60°C for 4 min.). Cycle sequenced products were resolved on an ABI 3730XL sequencer (Applied Biosystems, Foster City, CA, USA).

Table 1 325 adults of the *Urbanus bernikerni* complex reared from wild-caught caterpillars in Área de Conservación Guanacaste, caterpillar food plants and ecosystems, as of 31 Dec 2011

Food plant	<i>Urbanus signestami</i>	<i>Urbanus bernikerni</i>	<i>Urbanus ehakernae</i>
Asteraceae			
<i>Asteraceae 1454</i>	3	1	0
<i>Baltimora recta</i>	2	0	0
<i>Calea urticifolia</i>	10	2	0
<i>Clibadium leiocarpum</i>	3	1	2
<i>Clibadium pittieri</i>	18	4	8
<i>Clibadium surinamensis</i>	2	1	2
<i>Lasianthaea fruticosa</i>	1	0	0
<i>Lepidaploa tortuosa</i>	1	1	0
<i>Melanthera nivea</i>	75	43	3
<i>Otopappus verbesinoides</i>	3	0	0
<i>Salmea scandens</i>	5	1	0
<i>Verbesina ovatifolia</i>	6	8	0
<i>Viguiera dentata</i>	1	1	0
<i>Zexmenia virgulta</i>	75	23	19
Ecosystem	dry forest and rain forest	dry forest and rain forest	rain forest

EF1α sequences were heterogeneous, showing double peaks in sequence chromatograms. Consequently, polymorphic (heterogeneous) bases were scored using the IUPAC code for ambiguous bases. Sequences were edited using CodonCode Sequence V3.1.5 (CodonCode, Dedham, MA, USA) software, aligned using CLUSTAL W [26], and further manual inspection was done in MEGA 5.0 [27].

Sequences from *Urbanus proteus* (Linnaeus, 1758), *Urbanus esmeraldus* (Butler, 1877), *Urbanus evona* Evans, 1952,

and *Urbanus esta* Evans, 1952 were generated using the same protocols and were included as outgroups.

MrModelTest2.3 [28] was used to find the best nucleotide substitution model prior to Bayesian analysis. The GTR + G model was selected for both *COI* and *cytb*, while K80 + G was selected for *EF1α*. MrBayes v3.1.2 [29] was used to generate Bayesian gene trees for each marker. A total of 56 sequences were included in analysis of *COI*, 54 for *cytb*, and 30 for *EF1α*.

Sequencing success of the four amplified regions of *EF1α* was variable (Table 3). To maximize the number of specimens included in the Bayesian analysis, a sequence length of 541 bp was utilized. For quality assurance, sequence that had greater than 1% basepair ambiguities were excluded.

Four Monte Carlo Markov chains and a temperature of 0.2 were used for each analysis. Trees were sampled every 100 generations for 1,000,000 generations and Bayesian posterior probabilities were estimated for each node. The first 1250 trees (25%) were discarded as burn-in.

454-pyrosequencing analysis of *ITS2*

Thirty-nine individuals were selected for 454-pyrosequencing analysis. Multiple Identifier (MID) Tags were used to combine *ITS2* fragments from all individuals into a single 454 lane in a 16-lane sequencing run. *ITS2* fragments were PCR amplified with the MID tagged-primers using the same reaction mix mentioned above and *ITS2* was amplified under the following thermal cycling conditions: 94°C for 5 min., 40 cycles (94°C for 1 min., 53°C for 1 min., 72°C for 1 min.), 72°C for 10 min. All amplicons were sequenced on a 454 Genome Sequencer FLX System (Roche Diagnostics GmbH) following amplicon sequencing protocols (www.454.com). Amplicons of each sample were bidirectionally sequenced in 1/16th of a full sequencing run (70×75 picotiter plate).

Data was filtered using a 10-5-15 sliding window-phred score approach [30,31]. Sequences that passed this initial quality filtering underwent further filtering using the PRINSEQ webserver [32] to remove sequences with

Table 2 Description of PCR primers used in this study

Target	F primer	Primer sequence (5'-3')	R primer	Primer sequence (5'-3')	bp	Ref.
<i>COI</i>	LepF	ATTCACCAATCATAAAGATATTGG	LepR	TAAACTTCTGGATGTCCAAAAATCA	658	[12]
<i>cytb</i>	40 F	TGAATCTGAGGAGGGTTTGCIGT	560R	TCTACTGGCGGGCTCCAATTCA	499	
<i>EF1α</i>	Starsky	CACATYATTGTCGTSATYGG	Hutch	CTTGATGAAATCYCTGTGTCC	238	[23]
<i>EF1α</i>	Bo	GCTGAGCGYGARCGTGGTATCAC	Luke	CATRITGTCKCCGTGCCAKCC	367	[23]
<i>EF1α</i>	Laverne	GAGGAAATYAARAAGGAAG	Verdi	GACACCAGTTTCAACTCTGCC	267	[24]
<i>EF1α</i>	BJ	CARGACGTATACAAAATCGG	Bear	GCAATGTGRGCI GTGGCA	319	[24]
<i>ITS2</i>	<i>ITS2_d1F</i>	GAATGTCAGGACACATGAAC	<i>ITS4</i>	TCCTCCGCTTATTGATATGC	variable	[25]

Table 3 Sampling of *Urbanus bernikerni* complex specimens and gene region success

Voucher code	Identification	COI seq. length	n	cytb seq. length	n	EF1a seq. length	n/polymorphism	ITS2 seq. length
08-SRNP-4727	<i>Urbanus segnestami</i>	658	0n	499	0n	0	N/A	>200 bp
06-SRNP-47955	<i>Urbanus segnestami</i>	658	0n	499	0n	316	0n	>200 bp
06-SRNP-46570	<i>Urbanus segnestami</i>	657	0n	499	0n	1030	5n	>200 bp
06-SRNP-47304	<i>Urbanus segnestami</i>	632	0n	499	0n	316	2n	>200 bp
07-SRNP-41187	<i>Urbanus segnestami</i>	642	1n	499	0n	316	2n	>200 bp
08-SRNP-23800	<i>Urbanus segnestami</i>	658	0n	475	0n	315	2n	>200 bp
08-SRNP-5848	<i>Urbanus segnestami</i>	658	0n	456	0n	316	5n	>200 bp
08-SRNP-65935	<i>Urbanus segnestami</i>	658	0n	443	0n	1045	2n	>200 bp
08-SRNP-1439	<i>Urbanus segnestami</i>	658	0n	499	0n	1030	1n	>200 bp
08-SRNP-65153	<i>Urbanus segnestami</i>	658	0n	441	0n	1045	1n	0
07-SRNP-65582	<i>Urbanus segnestami</i>	658	0n	475	0n	1045	8n	>200 bp
07-SRNP-33232	<i>Urbanus segnestami</i>	658	0n	429	0n	1045	5n	>200 bp
07-SRNP-1119	<i>Urbanus segnestami</i>	658	0n	476	0n	1045	2n	0
07-SRNP-42421	<i>Urbanus segnestami</i>	658	0n	475	0n	1045	2n	0
07-SRNP-45090	<i>Urbanus segnestami</i>	658	0n	499	0n	933	7n	0
08-SRNP-24278	<i>Urbanus segnestami</i>	658	0n	0	N/A	0	N/A	>200 bp
06-SRNP-47801	<i>Urbanus bernikerni</i>	597	0n	499	0n	1030	0n	>200 bp
07-SRNP-30615	<i>Urbanus bernikerni</i>	658	0n	465	0n	541	3n	>200 bp
07-SRNP-20318	<i>Urbanus bernikerni</i>	658	0n	444	0n	0	N/A	>200 bp
06-SRNP-18276	<i>Urbanus bernikerni</i>	658	0n	499	0n	0	N/A	>200 bp
07-SRNP-57868	<i>Urbanus bernikerni</i>	658	0n	445	0n	541	0n	>200 bp
03-SRNP-12634.1	<i>Urbanus bernikerni</i>	658	0n	499	0n	0	N/A	0
06-SRNP-47816	<i>Urbanus bernikerni</i>	658	0n	499	0n	1030	2n	>200 bp
07-SRNP-45071	<i>Urbanus bernikerni</i>	655	0n	499	0n	316	0n	>200 bp
01-SRNP-4475	<i>Urbanus bernikerni</i>	658	0n	499	0n	316	1n	>200 bp
06-SRNP-3621	<i>Urbanus bernikerni</i>	658	0n	499	0n	0	N/A	0
07-SRNP-40856	<i>Urbanus bernikerni</i>	658	0n	499	0n	930	0n	0
07-SRNP-56547	<i>Urbanus bernikerni</i>	658	0n	443	0n	260	3n	>200 bp
07-SRNP-56829	<i>Urbanus bernikerni</i>	658	0n	446	0n	0	N/A	>200 bp
07-SRNP-30613	<i>Urbanus bernikerni</i>	658	0n	499	0n	541	3n	>200 bp
07-SRNP-40676	<i>Urbanus bernikerni</i>	658	0n	474	0n	541	1n	0
06-SRNP-47817	<i>Urbanus bernikerni</i>	658	0n	499	0n	0	N/A	0
06-SRNP-46460	<i>Urbanus bernikerni</i>	657	0n	499	0n	316	0n	>200 bp
06-SRNP-46639	<i>Urbanus bernikerni</i>	658	0n	499	0n	541	1n	0
06-SRNP-6511	<i>Urbanus bernikerni</i>	657	0n	499	0n	316	1n	0
06-SRNP-46572	<i>Urbanus bernikerni</i>	658	0n	499	0n	1030	3n	>200 bp
05-SRNP-1832	<i>Urbanus bernikerni</i>	658	0n	499	0n	0	N/A	0
02-SRNP-954	<i>Urbanus bernikerni</i>	658	0n	499	0n	316	1n	0
06-SRNP-65065	<i>Urbanus ehakernae</i>	649	0n	499	0n	1045	4n	0
07-SRNP-33182	<i>Urbanus ehakernae</i>	648	0n	499	0n	1045	3n	>200 bp
08-SRNP-65993	<i>Urbanus ehakernae</i>	658	0n	499	0n	1045	1n	>200 bp
08-SRNP-66028	<i>Urbanus ehakernae</i>	658	0n	499	0n	1045	4n	>200 bp
07-SRNP-33183	<i>Urbanus ehakernae</i>	658	0n	499	0n	541	0n	>200 bp
07-SRNP-33184	<i>Urbanus ehakernae</i>	658	0n	499	0n	1045	5n	>200 bp

Table 3 Sampling of *Urbanus bernikerni* complex specimens and gene region success (Continued)

07-SRNP-287	<i>Urbanus ehakernae</i>	658	0n	499	0n	316	0n	0
07-SRNP-40761	<i>Urbanus ehakernae</i>	655	0n	499	0n	1045	1n	>200 bp
06-SRNP-46887	<i>Urbanus ehakernae</i>	657	0n	499	0n	316	5n	>200 bp
06-SRNP-46891	<i>Urbanus ehakernae</i>	657	0n	499	0n	1030	11n	>200 bp
05-SRNP-40545	<i>Urbanus ehakernae</i>	658	0n	499	0n	541	0n	>200 bp
05-SRNP-41753	<i>Urbanus ehakernae</i>	658	0n	499	0n	316	3n	>200 bp
06-SRNP-43129	<i>Urbanus ehakernae</i>	381	0n	499	0n	1030	2n	>200 bp
07-SRNP-57988	<i>Urbanus evona</i>	658	0n	499	0n	1030	2n	0
05-SRNP-66411	<i>Urbanus evona</i>	658	0n	499	0n	851	0n	>200 bp
04-SRNP-15286	<i>Urbanus evona</i>	658	0n	498	1n	848	3n	0
07-SRNP-1369	<i>Urbanus proteus</i>	658	0n	499	0n	314	2n	0
08-SRNP-71995	<i>Urbanus proteus</i>	658	0n	499	0n	415	4n	0
08-SRNP-21596	<i>Urbanus esmeraldus</i>	658	0n	499	0n	313	3n	0
07-SRNP-57862	<i>Urbanus esmeraldus</i>	658	0n	499	0n	315	1n	0
07-SRNP-57861	<i>Urbanus esmeraldus</i>	658	0n	0	N/A	416	3n	0
06-SRNP-1984	<i>Urbanus esta</i>	658	0n	443	0n	845	6n	0
07-SRNP-56430	<i>Urbanus esta</i>	658	0n	499	0n	848	3n	0
07-SRNP-57988	<i>Urbanus evona</i>	658	0n	499	0n	1030	2n	0

Sequence length and number of ambiguous bases and/or polymorphic positions [n] are included for each specimen.

base ambiguities. A sequence length threshold of 200 bp was selected to preserve nucleotide variation and the number of sequences included in the analysis. Chimera detection was performed in UCHIME using the *de novo* option [33]. Forward and reverse direction sequences were reduced to variants by collapsing sequences that were 100% identical to each other, using the CD-HIT webserver [34]. As the final quality control, to further reduce potential impact of 454-pyrosequencing error, we only included variants that had 2 or more collapsed sequences. Only forward direction *ITS2* sequences were used in ordination analyses (see below) because reverse variants, after filtering, were minimal (1 to 3) for each individual. Forward direction variants of each individual were aligned in MEGA 5.0 using ClustalW with an alignment gap opening value of 7 and a gap extension value of 3, followed by manual editing.

Ordination by non-metric multidimensional scaling (nMDS), using the Bray-Curtis similarity measure [35] was performed in PRIMER v6 [36] to create a graphical representation of the intragenomic, intraspecific, and interspecific clustering of *ITS2*. A similarity matrix of variants found within each individual was created by treating every position in the sequence alignment as a separate character, and nucleotides were coded arbitrarily as A = 0, C = 1, G = 2, T = 3 [7] and gaps were treated as a fifth character, gap = 4, in the matrix.

We assembled forward and reverse direction variants of *ITS2* that had at least 95% match in a 20 bp overlap. These full-length *ITS2* sequences were aligned in MEGA

5.0 and maximum parsimony (MP) trees were generated in PAUP 4b1.0 [37] with and without gaps as a fifth base. A heuristic search was performed with tree bisection reconnection (TBR) branch swapping and 1,000 random taxon addition replicates, saving no more than 100 equally parsimonious trees per replicate. One thousand bootstrap replicates were performed and 50% majority rule trees were created. Nodes were collapsed below the 50% confidence interval and only bootstrap values above 50% confidence interval were included.

***ITS2* secondary structure and CBC analysis**

Full-length *ITS2* intragenomic variants were collapsed by 100% identity to reduce redundancy of the dataset using the CD-HIT web-server [34]. Compensatory base changes (CBCs) occur when both nucleotides of a paired site mutate while the pairing itself is maintained (G-C to A-U) [38]. Secondary structure was determined for unique *ITS2* variants using the Vienna RNAfold Webserver (<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>). A default setting of a folding temperature at 37°C, allowing for dangling energies on both sides of a helix, using minimum free energy and partition function fold algorithms was employed. The program 4SALE [39] was used to align secondary structure and estimate CBCs between variants.

***Wolbachia* assays**

The presence of *Wolbachia* infection was established by a PCR test for the *Wolbachia* surface protein (*wsp*) in leg tissue. PCR reaction mixture and amplification protocols

for the *wsp* marker are given by Baldo et al. [18]. If *wsp* PCR bands were present in some or all individuals of a DNA barcode lineage, all specimens from that lineage underwent the more extensive Multilocus Sequence Typing (MLST) assay [18] to identify the specific strain of *Wolbachia*. PCR band checks, Sanger sequencing protocols, and editing are the same as described for *COI* and *cytb*.

Results

Description of cryptic species

Permanent names are needed for further discussion of the three cryptic species in this study. As noted above, all three species are passing under the one name *Urbanus belli*, a species currently treated as pan-neotropical although it was described only from northern Argentina [40]. Given three cryptic species in northwestern Costa Rica, the different geologic history of Meso- and South America, and the degree to which we are finding that supposedly single species occurring in both continents actually comprise two or more related species that are much more limited in distribution (often replacing one another geographically), we believe that extending the application of the name *Urbanus belli* to ACG lowland Asteraceae-eating *Urbanus* is unwarranted.

Evans [41] correctly judged that true *Urbanus belli* is closely related to *Urbanus viterboana* [42], which in ACG is the high elevation Asteraceae-eating *Urbanus*. He went so far as to treat *Urbanus belli* as a subspecies of it with a limited, far southern distribution, recording *Urbanus viterboana belli* only from Argentina and Bolivia. However, he also described a new subspecies, *Urbanus viterboana alva*, which he recorded as pan-neotropical (Mexico to Argentina). Were *alva* an applicable epithet, it would pertain to a full species because it appears to be broadly sympatric with *Urbanus viterboana viterboana*, which Evans likewise recorded as pan-neotropical (Mexico to Bolivia). Morphology of the intricate genitalia of skippers is often diagnostic of species. Evans [41: 85–87; plate 18, figure C.13.2] described *Urbanus viterboana alva* from both the “very variable” superficial appearance of the adult and the male genitalia (which are somewhat variable but, at the same time, morphologically simple as skipper genitalia are concerned). Steinhäuser [43] pointed out that the Evans dissection of the holotype of *Urbanus viterboana alva* is partial and in poor condition, and that the broken genitalic parts (comprising only the claspers) on the associated card cannot possibly have come from the holotype, so that one cannot say to what taxon the name should apply. We are disregarding the epithet *alva*.

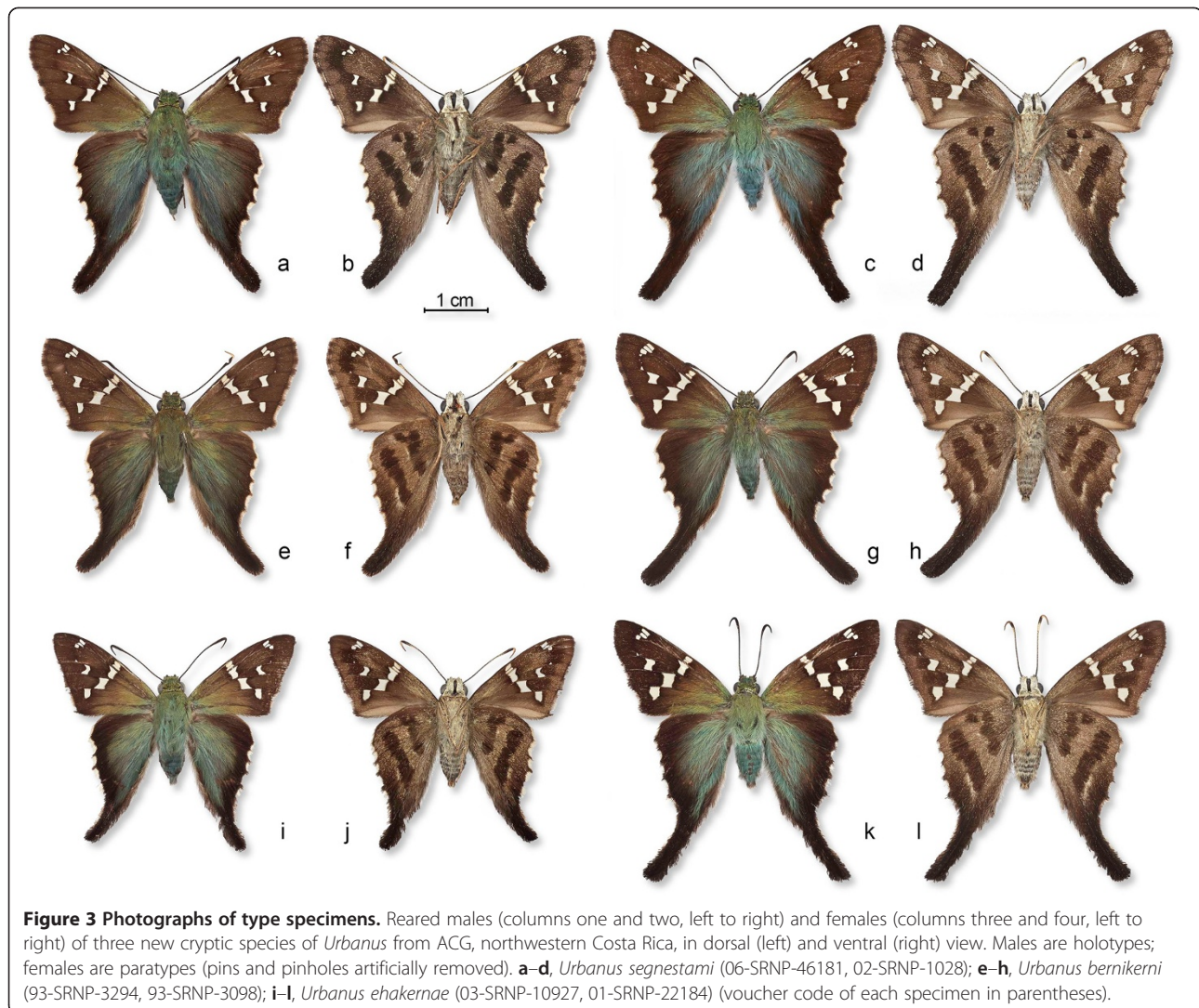
In the Neighbour-Joining trees based on DNA barcodes of ACG skippers (Figure SI in [11]), two of the three cryptic species cluster more closely with the species *Urbanus viterboana* than with the third cryptic species.

We call this four-species cluster the *Urbanus viterboana* group. Natural history evidence strongly supports this grouping: so far as is known (and amply documented in the ACG *Urbanus* caterpillar inventory [21]), caterpillars of the species of *Urbanus* that feed on dicots—and caterpillars of the species in many of its related genera as well—feed almost exclusively on legumes (Fabaceae), except for the four species of the *Urbanus viterboana* group, which eat the mature leaves of herbaceous composites (Asteraceae). These large plant families are not phylogenetically close to each other, and the skippers’ evolutionary shift in larval foodplant selection underlines their tight relationship and the distinctiveness of the *Urbanus viterboana* group.

The three cryptic species initially revealed by barcodes are superficially indistinguishable (Figure 3) and *Urbanus viterboana* itself is not much different (see images in [21]). Although the uppersides of their wings differ greatly from the undersides, this difference is similar across all three species (Figure 3). Hindwing tails are longer in females (Figure 3c, d, g, h, k, l) than in males (Figure 3a, b, e, f, i, j), but both sexes express the same slight variations in color pattern. The most evident variation is at the top end of the innermost of the two dark brown bands on the ventral hindwing, where the phenotype ranges from two separate spots (Figure 3d) to variable degrees of union with one another and with the band itself. This trio of cryptic species is here designated the *Urbanus bernikerni* complex of the *Urbanus viterboana* group. Because the structural color on the upperside of its wings is much more blue than green, *Urbanus viterboana* is superficially separable from the species of the *Urbanus bernikerni* complex (at least when specimens are in good condition). Likewise, the hindwing tails of *Urbanus viterboana* are slightly longer than are those of the *Urbanus bernikerni* complex. *Urbanus viterboana* lives at higher elevations (at least 900 m) than do the cryptic species.

Probing the possibility that species of the *Urbanus bernikerni* complex differ in communicative characters, such as ultraviolet reflectance patterns of the wings or pheromones released by a secondary sex character on the forewing of males, was beyond the scope of our study.

The following original taxonomic descriptions of these look-alike species are necessarily brief (Table 4). Except for the barcodes, the molecular characters and the analysis and interpretation of them are too complex to express concisely. The characters require discussion, and they have their limitations (see subsequent sections of this paper). Speciation has no regard for taxonomists. However, the DNA barcodes themselves are highly conservative within each species of the *Urbanus viterboana* group and significantly different among them (Figure 1 and appendix to [11]). The holotype chosen for each new species has the predominant barcode of that species; and the barcode,



comprising 658 base pairs in each case, is given. The voucher code for a specimen leads to detailed information about the stage in which that specimen was collected, what it was eating, exactly where it was found, etc.—information that is provided in the ACG inventory database [21].

As required by the International Code of Zoological Nomenclature for electronic publication of new species, this work is registered in ZooBank as (urn:lsid:zoobank.org:pub:B033EAF0-C9C9-47 F1-B4D8-FAA829F6876B), and will be electronically archived by PubMed Central (<http://www.ncbi.nlm.nih.gov/pmc>) once published. The date of publication will be the date of publication of this paper by BMC Evolutionary Biology (ISSN 1471–2148).

Phylogenetic analysis of *COI*, *cytb*, and *EF1 α*

Sequencing success for each marker and all specimens is shown in Table 3. Sequences were also submitted to the Dryad repository and can be found at doi:10.5061/dryad.pj561. The results of the Bayesian analyses of *COI*, *cytb*,

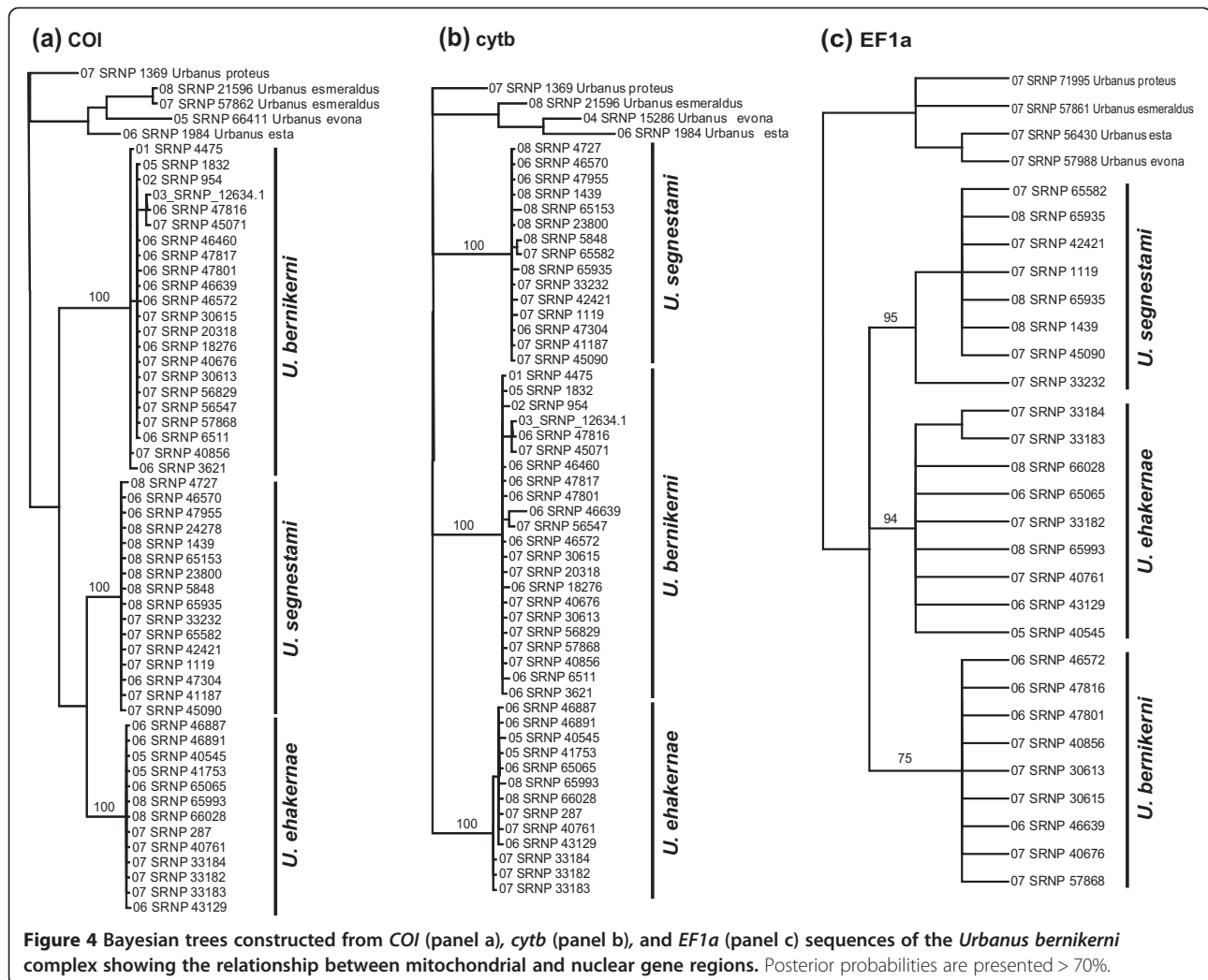
and *EF1 α* show three supported clades (Figure 4). These clades are congruent with the species boundaries suggested by DNA barcode data (Figure 1).

Analysis of *ITS2* sequences

The 39 individuals successfully 454-pyrosequenced for *ITS2* produced 5,315 sequence reads, which after filtering consisted of 303 variants (Table 5). The nMDS plot (Figure 5) shows distinct separation of *Urbanus bernikerni* from *Urbanus segnestami* and *Urbanus ehakernae*. Two clusters are observed for *Urbanus segnestami*, which highlights the level of apparently intraspecific variability that is found within this lineage, though the possibility of yet more phylogenetic structure in this population cannot be unambiguously discarded. Two individuals, 06-SRNP-46887 and 06-SRNP-46891 of *Urbanus ehakernae*, each possess a variant (06-SRNP-46887_a, 06-SRNP-46891_d) that falls into the *Urbanus segnestami* cluster. These two individuals are most likely siblings. Both were found at the

Table 4 Descriptions of the three new species comprising the *Urbanus bernikerni* complex, based on specimens reared from wild-caught caterpillars in ACG, northwestern Costa Rica, and deposited in USNM

Characters	<i>Urbanus segnestami</i> Burns n. sp.	<i>Urbanus bernikerni</i> Burns n. sp.	<i>Urbanus ehakernae</i> Burns n. sp.
Holotype male voucher code	06-SRNP-46181	93-SRNP-3294	03-SRNP-10927
Holotype DNA barcode & Genbank Accession No.	AACTTTATATTTTATTTTTGGAATT TGAGCAGGATTAATTGGAACCTT CTCTAAGATTACTTATTCGAACT GAATTAGGAACCCAGGATCTTT AATTGGAGATGATCAAATTTATA ATACTATTGTAACAGCTCACGCA TTTATTATAATTTTCTTTATAGTT ATACCTATTATAAATGGAGGATT TGGTAATTGACTAGTACCTTTAA TAATAGGAGCCCTGATATAGCT TTCCCCGTATAACAATATAAG ATTTTGATTACTACCCCATCTTT AACTTTATAATTTCAAGAAGAAT TGTTGAAAATGGTGTCTGACTG GATGAACAGTTTATCCCCCTTT CATCTAATATTGCCATCAAGGAG CTTCTGTTGATTTAGCAATTTTTT CTTCATCTTGCAGGTATTTATC AATTCTTGAGCTATTAATTTTAT TACAACAATTATTAATACGAAT TAATAGATTATCTTTTATCAAAAT ACCATTATTTGTATGAGCTGTAG GAATTACAGCATTATTTATTTA CTTTCTTACCTGTTTAGCTGG AGCTATTACTATATTTAACTG ACCGAACTTAAATACTTCATTT TTTGATCCTGCTGGAGGAGGAG ATCCAATTCTATATCAACATTTA TTT, GU156383	AACCTTATATTTTATTTTTGGAAT TTGAGCAGGATTAATTGGAACCT CTTTAAGATTACTTATTCGAACT GAATTAGGAACCCAGGATCTCT AATTGGAGATGATCAAATTTATA ATACTATTGTAACAGCTCATGCA TTTATTATAATTTTCTTTATAGTT ATACCTATTATAAATGGAGGATT TGGTAATTGATTAGTCCCTTTAA TAATAGGTGCTCTGATATAGCT TTCCCCGTATAACAATATAAG ATTTTGATTACTACCCCATCTTT AACTTTATAATTTCAAGAAGAAT TGTTGAAAATGGAGCTGACTG GATGAACAGTTTATCCCCCTTT TCATCTAATATTGCTACCAAGG AGCCTCAGTTGATTTAGCAATTT TTCCCTTCATCTGCTGCTGATTT CATCAATTCTGGAGCTATTAAT TTTATTACAACAATTATTAATAT ACGAATTAATAGATTAACTTTTG ATCAAATACCTTTATTTGTATGA GCCGTAGGAATTACAGCATTATT ATTATTACTTTCTTACCTGTTT TAGCTGGAGCTATTACTATTTA TTAACAGATCGAAATTTAAATAC ATCGTTTTTGTATCCTGCTGGAG GAGGAGATCCAATTTTATACCAA CATTATTT, JQ526818	AACTTTATATTTTATTTTTGGAAT TTTGAGCAGGATTAATTGGAACCT TCCTTAAGATTACTTATTCGAACT GAATTAGGAACCCAGGATCTTT AATTGGAGATGATCAAATTTATA ATACTATTGTTACAGCTCACGCAT TTATCATAATTTTCTTTATAGTTA TACCTATTATAAATGGAGGATT GGTAATTGATTAGTACCTTTAAT AATAGGAGCCCTGATATAGCTT TCCCTCGTATAACAATATAAGA TTTTGATTACTCCTCTCTTT AACTTTATAATTTCAAGAAGAA TTGTTGAAAATGGTGTCTGACT GGATGAACAGTTTATCCCCCTCT TCATCTAATATTGCCACCAAG GAGCTTCTGTTGATTTAGCAATTT TTTTCTTCTATAGCAGGTAT CTCATCAATTCTGGAGCTATTA ATTTTATTACAACAATCATTAA TACGAATTAATAGATTATCTTTTG ATCAAATACCATTATTTGTGTGAG CTGTAGGAATTACAGCATTATTT TATTACTTTCTTACCTGTTTAC TGGAGCTATTACTATTTAACTT GATCGAACTTAAATACTTCATTT TTGACCTCTGCTGGAGGAGGAGAT CCAATTCTATATCAACATTTATT GU161980
Paratypes	107 males, 84 females	39 males, 36 females	16 males, 16 females
DNA barcode cluster in NJ tree	See Figure 1	See Figure 1	See Figure 1
Additional molecular characters	See subsequent text	See subsequent text	See subsequent text
Ecosystem	Dry forest and rain forest	Dry forest and rain forest	Rain forest
Presence of <i>Wolbachia</i> strain 108 of Supergroup B	Yes	No	Yes
Etymology	Named in honor of Mats Segnestam of Sigtuna, Sweden, formerly of Svenska Naturforeningen and then the Swedish governmental agency SIDA, in recognition of his intense two decades of facilitating the germination and growth of ACG as a concept and reality of conservation through non-damaging biodiversity development.	Named in honor of Bernd Kern (RIP) of Vaesterhaninge, Sweden, in recognition of his joint high energy, dedication, and enthusiasm with Eha Kern in the 1980's creation and subsequent decades of execution of Barnensregnskog (Childrens Rainforest of Sweden, http://www.barnensregnskog.se) in support of land purchase for ACG and support of many other Neotropical conservation projects.	Named in honor of Eha Kern of Vaesterhaninge, Sweden, in recognition of her joint high energy, dedication, and enthusiasm with Bernd Kern in the 1980's creation and subsequent decades of execution of Barnensregnskog (Childrens Rainforest of Sweden, http://www.barnensregnskog.se) in support of land purchase for ACG and support of many other Neotropical conservation projects.



very western-most edge of the distribution of *Urbanus ehakernae* as it protrudes into the more encompassing “dry forest to rain forest” distribution of *Urbanus segnestami*.

The alignment of *ITS2* sequences in the MP analyses with gaps treated as missing data had 28 parsimony-informative characters (Figure 6a), while the *ITS2* alignment with gaps treated as a fifth character had 41 parsimony-informative characters (Figure 6b). The MP analysis of the *ITS2* alignment that included gaps as fifth characters (CI = 0.853, RI 0.937) resulted in 144 most-parsimonious trees; and the 50% majority-rule consensus tree supports reciprocal monophyly of *Urbanus bernikerni*, but shows *Urbanus ehakernae* to be paraphyletic to *Urbanus segnestami* (Figure 6b). The paraphyly is caused by the 2 variants mentioned above, 06-SRNP-46887_a and 06-SRNP-46891_d (Figure 6b). If gaps are not considered as a fifth base, the MP analysis resulted in 770 most-parsimonious trees (CI = 0.861, RI = 0.954); and the 50% majority-rule consensus tree shows less than 50% support for the *Urbanus segnestami* node (Figure 6a).

ITS2 secondary structure & compensatory base changes

The secondary structures of *Urbanus bernikerni* complex variants show the conserved *ITS2* structure usually found in eukaryotes, including four helices (labeled I-IV, Figure 7) with helix III being the longest, containing a 5' end GGU motif, as well as the U-U mismatch associated with the *ITS2* helix II in most eukaryotes [44]. The CBC analysis identified 10–14 compensatory base pair changes among the three *Urbanus bernikerni* complex lineages in helices II and III (Table 6). Intraspecific and intragenomic CBCs were found in *Urbanus ehakernae*. CBCs were found in helix III between sequence type 06-SRNP-46891_a and four other variants: 06-SRNP-46887_c, 05-SRNP-41753, 05-SRNP-42394, 06-SRNP-46891_d (Table 6). This variant (06-SRNP-46891_a), also clusters within the *Urbanus segnestami* clade in the MP trees (Figures 6a, b).

Wolbachia assays

A total of 16 individuals were positive for the *wsp* and MLST markers. This included 40% (4/10) of individuals

Table 5 Results of 454-pyrosequencing *ITS2* from the *Urbanus bernikerni* complex

MID	Sample ID	Species	No. sequences	No. unique variants	No. variants (>2 seq)	No. sequences in variants (>2 seq)
22	07-SRNP-30614	<i>Urbanus segnestami</i>	133	117	7	16
24	07-SRNP-33232	<i>Urbanus segnestami</i>	147	136	6	11
26	08-SRNP-1439	<i>Urbanus segnestami</i>	207	172	15	35
27	08-SRNP-65935	<i>Urbanus segnestami</i>	66	46	8	20
28	08-SRNP-5848	<i>Urbanus segnestami</i>	174	150	11	24
29	08-SRNP-23800	<i>Urbanus segnestami</i>	142	117	9	25
30	08-SRNP-24278	<i>Urbanus segnestami</i>	183	137	18	46
52	07-SRNP-41187	<i>Urbanus segnestami</i>	120	91	13	29
53	06-SRNP-47304	<i>Urbanus segnestami</i>	80	63	8	17
54	06-SRNP-47955	<i>Urbanus segnestami</i>	170	117	23	53
55	06-SRNP-46570	<i>Urbanus segnestami</i>	143	115	13	28
56	08-SRNP-4727	<i>Urbanus segnestami</i>	43	36	3	7
31	06-SRNP-18276	<i>Urbanus bernikerni</i>	106	90	8	16
32	06-SRNP-46572	<i>Urbanus bernikerni</i>	151	120	24	71
33	07-SRNP-30615	<i>Urbanus bernikerni</i>	179	132	20	47
34	07-SRNP-30613	<i>Urbanus bernikerni</i>	233	189	17	44
37	07-SRNP-20318	<i>Urbanus bernikerni</i>	30	24	3	6
38	07-SRNP-57868	<i>Urbanus bernikerni</i>	56	50	3	6
39	07-SRNP-56547	<i>Urbanus bernikerni</i>	82	73	4	9
40	07-SRNP-56829	<i>Urbanus bernikerni</i>	322	266	23	56
57	07-SRNP-45071	<i>Urbanus bernikerni</i>	129	109	8	10
64	06-SRNP-47801	<i>Urbanus bernikerni</i>	240	166	30	74
65	06-SRNP-46460	<i>Urbanus bernikerni</i>	147	115	12	32
68	05-SRNP-1832	<i>Urbanus bernikerni</i>	4	4	0	0
69	01-SRNP-4475	<i>Urbanus bernikerni</i>	74	68	3	6
41	05-SRNP-41753	<i>Urbanus ehakernae</i>	164	125	17	39
42	05-SRNP-40545	<i>Urbanus ehakernae</i>	108	88	7	20
43	06-SRNP-65065	<i>Urbanus ehakernae</i>	11	11	0	0
45	07-SRNP-40761	<i>Urbanus ehakernae</i>	144	93	21	51
46	07-SRNP-33184	<i>Urbanus ehakernae</i>	169	118	18	51
47	07-SRNP-33183	<i>Urbanus ehakernae</i>	101	74	11	27
48	07-SRNP-33182	<i>Urbanus ehakernae</i>	117	80	17	37
49	08-SRNP-66028	<i>Urbanus ehakernae</i>	106	76	10	30
50	08-SRNP-65993	<i>Urbanus ehakernae</i>	156	126	13	30
70	06-SRNP-43129	<i>Urbanus ehakernae</i>	134	91	17	34
71	06-SRNP-46891	<i>Urbanus ehakernae</i>	190	156	11	34
72	06-SRNP-46887	<i>Urbanus ehakernae</i>	293	237	26	56
73	05-SRNP-42394	<i>Urbanus ehakernae</i>	112	87	11	25

of *Urbanus segnestami*, 100% (12/12) of the individuals of *Urbanus ehakernae*, and 0% (0/13) of *Urbanus bernikerni*. We sequenced all PCR bands for *wsp* and MLST markers. Sequences generated by extending the reverse primers for *gatB*, *coxA*, *ftsZ*, *fbpA*, and *wsp* showed

heterogeneous sequence chromatograms for both *Urbanus segnestami* and *Urbanus ehakernae* specimens. The *hcpA* marker showed heterogeneous sequence chromatograms for both forward and reverse primer sequences. In all cases, the homogeneous sequence read or the highest

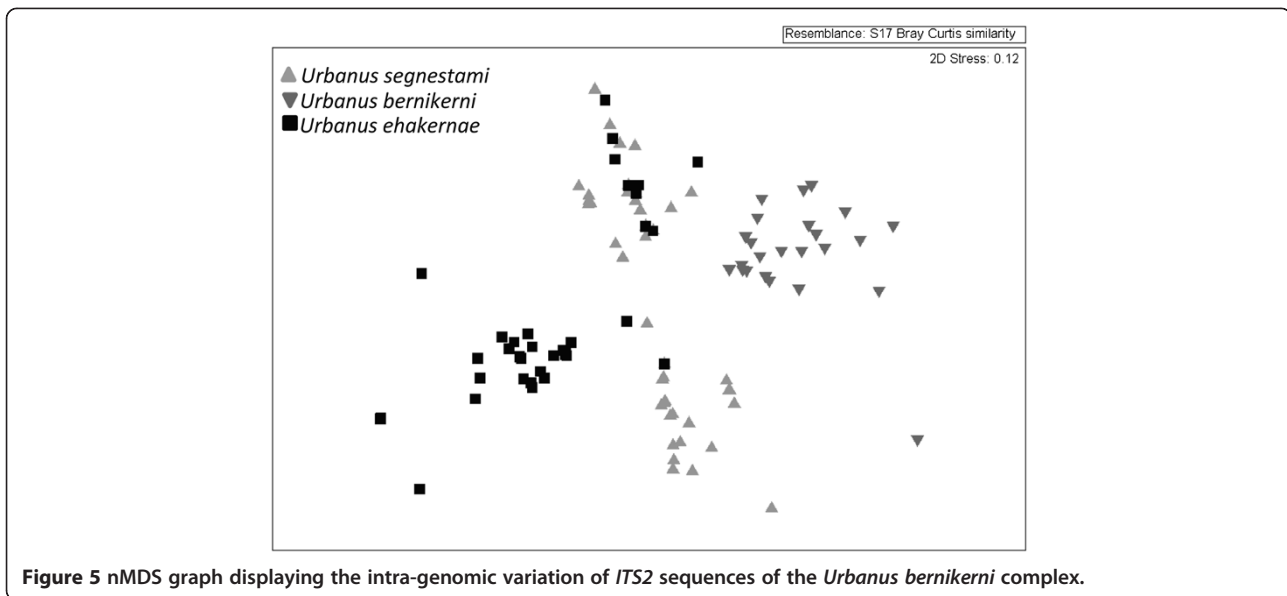


Figure 5 nMDS graph displaying the intra-genomic variation of *ITS2* sequences of the *Urbanus bernikerni* complex.

peak was used as the consensus sequence to search the *wsp* and MLST database. All six loci matched known allelic profiles with 98-100% identity, and all infections are therefore *Wolbachia* strain 108 of Supergroup B (Table 7). Strain 108 was previously found in a butterfly, *Brangas felderi* (Goodson, 1945) (Lycaenidae), in Morona-Santiago, Ecuador [45,46]; genus *Brangas*, but not this species, does occur in ACG, though not in ecological proximity to the *Urbanus viterboana* complex.

Discussion

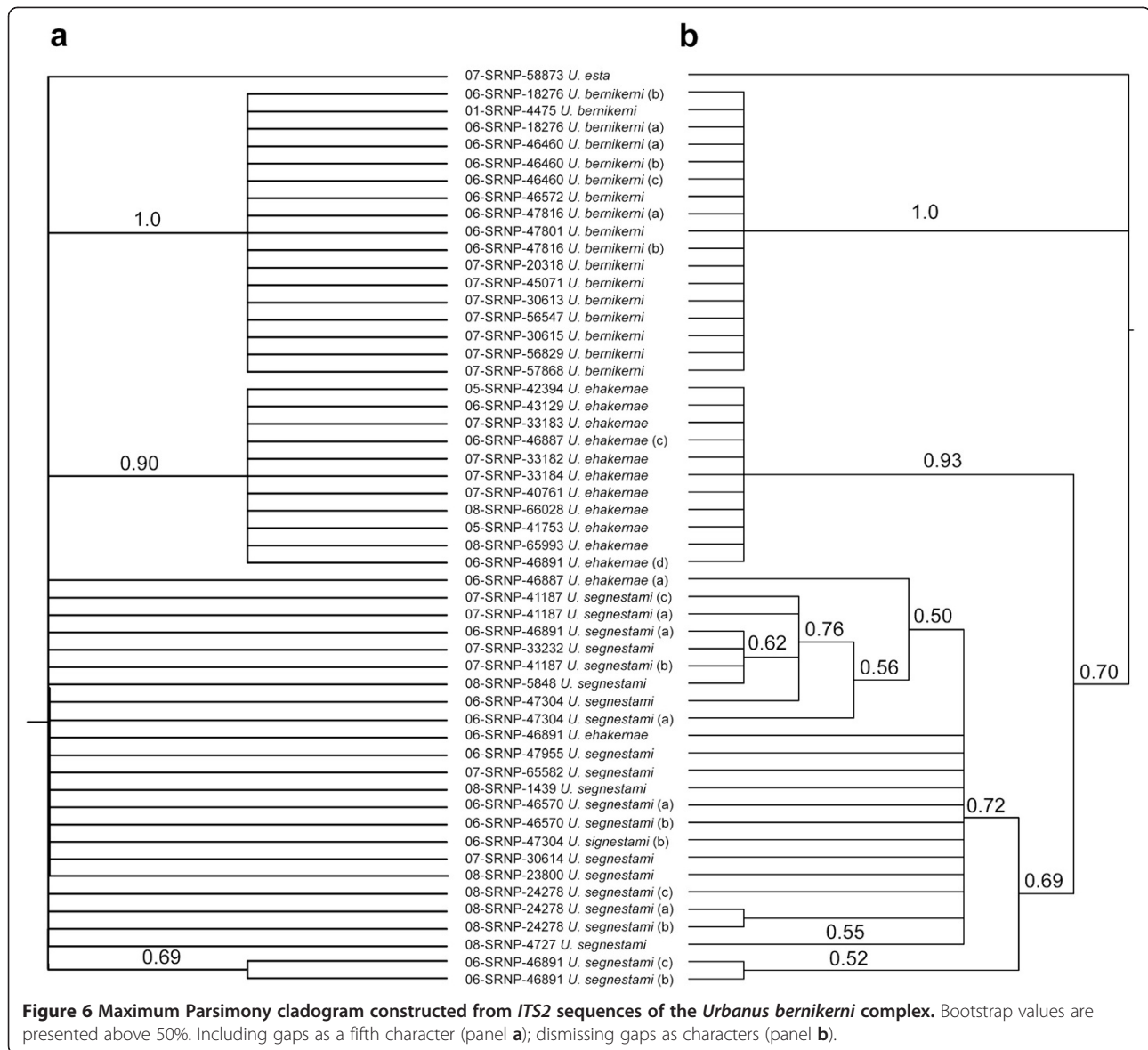
This study provides genetic evidence to support the separation of what was previously called “*Urbanus belli*”, found in ACG lowland rain forest and dry forest, into a morphologically cryptic complex of three species. We found congruent patterns across 4 loci in support of three genetically distinct lineages. The fact that our cryptic complex occurs in sympatry provides indirect evidence that each lineage, when given the opportunity, does not appear to exchange genetic material. Although the intragenomic variability of *ITS2* of two individuals of *Urbanus ehakernae* could be the result of hybridization or the retention of shared ancestral polymorphisms, the ancillary ecological data collected in this study supports evolutionary distinction. *Urbanus ehakernae* is the only species of the complex restricted to rainforest habitat and this lineage appears to maintain its evolutionary trajectory regardless of gene exchange.

This study successfully implements a genetic framework for cryptic species detection and characterization. The remaining sections of this discussion focus on the application and limitations of such a framework.

ITS2 as a species level marker

ITS2 is a segment of the nuclear ribosomal DNA (rDNA) cistron and has been used extensively for DNA barcoding analyses, especially in plants and fungi [47,48]. The number, size, and location of rDNA arrays can vary within species and individuals [49] and, as a result, influence the utility of the marker for phylogenetic interpretation at the species level. The process of concerted evolution homogenizes rDNA arrays and theoretically removes variability among repeats within the genome. However, intragenomic variants appear to be common [15,50,51]. For example, intragenomic variation of *ITS2* was characterized from 178 plant species [15]. The study showed an average of 35 variants per species and detected cases of shared variants in congeneric species [15]. Despite the presence of shared variants, the overall intragenomic variability was negligible, as differences were markedly smaller than those of intraspecific and interspecific differences, yielding a 97% success rate for species level identification [15]. The findings of our study agree with that of [15]; we found few shared variants between species. Although we provide inconclusive evidence as to the nature of these variants, their existence can shed light on the evolutionary history of the species complex. Similarly, studies utilizing internal transcribed spacer region 1 (*ITS1*) to confirm *COI* lineages of provisional cryptic species of parasitoid flies (Diptera, Tachinidae) and Malagasy ants [4,52,53] have also provided inconclusive evidence to address the cause of *ITS2* heterogeneity. Future work that includes closely related congeneric species over a broader geographical distribution would help to elucidate these findings.

Previous studies have utilized cloning experiments and 454-pyrosequencing to separate heterogeneous *ITS2*



sequences [15,54]; the latter is analogous to a massive-scale cloning experiment. The benefits of using 454-pyrosequencing are the massive number of sequences acquired and the ability to target major and minor variants of *ITS2* [15]. More than 5000 sequences, 200–400 bp in length, were obtained after filtering and resulted in documenting over 300 variants of *ITS2*. However, a major drawback to the 454-Roche platform is the susceptibility of over/under and delayed base calling, either of which can lead to sequencing inaccuracies [55]. These inaccuracies are difficult to separate from the tendency of *ITS2*—as a non-coding region—to accumulate indels. We used a rigorous filtering process to help eliminate 454-pyrosequencing errors. Additionally, we only included variants with two or more identical sequences in the analyses, which should

ensure exclusion of random sequence errors. In addition, sequencing success was low from the reverse primer direction. We suspect this is due to the preferential sequencing of primer dimers or to unperceived secondary structure in *ITS2* sequences. This limited the interpretation of ~250 bp sequences in the forward direction for intragenomic variability, and reduced the number of full-length *ITS2* variants that could be used in the MP and CBC analysis. Despite these limitations, the nMDS results suggest that the ~250 bp sequences adequately illustrate the level of genetic variability of *ITS2* reflected in the MP analysis. The quality of *ITS2* sequences recovered from NGS platforms will undoubtedly improve as NGS technology advances, ultimately increasing the utility of multi-copy genes, like *ITS2*, for phylogenetic analysis.

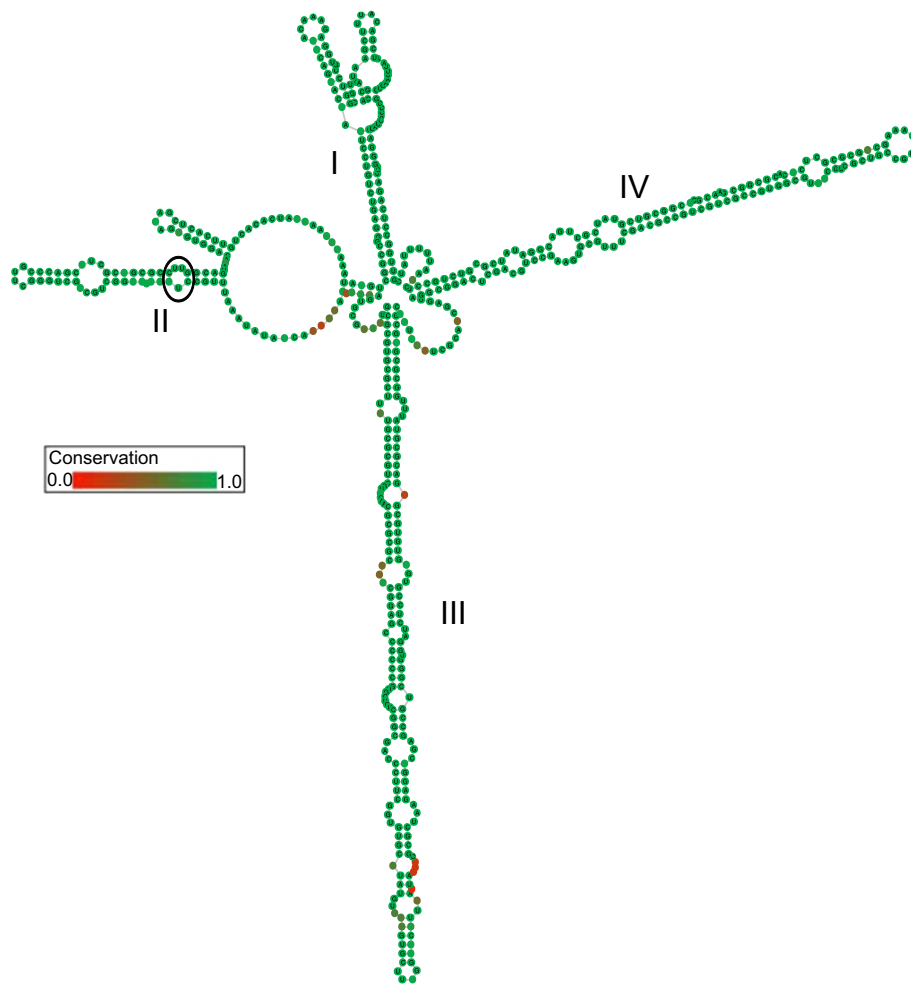


Figure 7 Conserved *ITS2* secondary structure for variants of the *Urbanus bernikerni* complex. Helices are numbered accordingly. The U-U mismatch typically found in Eukaryotes is outlined with a black circle in helix II. The complete structure represents the 51% consensus of aligned structures with gaps. Degree of secondary structure conservation is displayed in color grades from green (conserved) to red (not conserved).

Compensatory base changes in *ITS2* secondary structure

CBCs are crucial to preserve the pairings required for helices that make up the core of the *ITS2* secondary structure [56]. CBCs most likely have little causal relationship to speciation, but they may indicate that adequate evolutionary time has elapsed for speciation to

Table 6 CBC matrix showing the number of compensatory base changes within and between each provisional species

	<i>Urbanus segnestami</i>	<i>Urbanus bernikerni</i>	<i>Urbanus ehakernae</i>
<i>Urbanus segnestami</i>	0		
<i>Urbanus bernikerni</i>	17	0	
<i>Urbanus ehakernae</i>	10	12	4

Table 7 *Wolbachia* BLAST results from the *wsp* and *MLST* database for the five markers: *coxA*, *hcpA*, *gatB*, *ftsZ*, and *fbpA*

Marker	Allele	Percentage identity match in WSP & MLST databases
<i>wsp</i>	115	100%
<i>gatB</i>	71	98%
<i>coxA</i>	67	100%
<i>ftsZ</i>	65	99%
<i>hcpA</i>	74	99%
<i>fbpA</i>	6	98%

The table includes the allele type that was found in the BLAST search and the percentage identity match in the databases resulting in the *Wolbachia* strain 108 of Supergroup B.

have occurred [16,57]. The numbers of CBCs in *ITS2* secondary structure helices have been shown to correlate with closely related taxa being reproductively incompatible and have identified biological species of plants 93.11% of the time [15-17,56]. Previous findings that investigated the effects of *ITS2* intragenomic variability on the efficacy of CBCs to identify species found that 97% of 178 plant species could be identified, when including the relative abundance of variants [17]. The authors do state that identical *ITS2* variants were found across intra- and intergeneric species, and in a related study, 17% of the genera had congeneric species that shared *ITS2* variants [15,17]. Likewise, in our study, 39 interspecific CBCs were found between members of the *Urbanus bernikerni* complex, providing further evidence to validate this new classification (Table 6). Nevertheless, 2 intragenomic and 2 intraspecific CBCs were found within and between 2 individuals of *Urbanus ehakernae*, and other members of the lineage. The *ITS2* profiles of these two individuals are heterogeneous, including variants characteristic of both *Urbanus ehakernae* and *Urbanus segnestami*. As such, it is not surprising to find intragenomic and intraspecific CBCs within *Urbanus ehakernae*. As discussed in the previous section, future work focused on elucidating the nature of intragenomic and intraspecific *ITS2* variants can shed light on the evolutionary history of this cryptic complex.

Wolbachia endosymbiosis

Urbanus segnestami and *Urbanus ehakernae* were found to share the same *Wolbachia* strain, while infection was completely absent in *Urbanus bernikerni*. This sharing between *Urbanus segnestami* and *Urbanus ehakernae* can be the result of two phenomena. The ancestor to *Urbanus segnestami* and *Urbanus ehakernae* could have been infected with this strain of *Wolbachia* and today we are observing a secondary loss of the infection in *Urbanus segnestami*, and fixation of the infection in *Urbanus ehakernae*. On the other hand, this sharing can represent unidirectional introgression of *Wolbachia* from *Urbanus ehakernae* to *Urbanus segnestami* and signify interbreeding of the lineages upon secondary contact in ACG (horizontal gene transfer).

The presence of *Wolbachia* in *Urbanus segnestami* and *Urbanus ehakernae*, and absence of infection in *Urbanus bernikerni*, can suggest that unidirectional cytoplasmic incompatibility (uniCI) [19] is a mechanism that is reducing gene flow between these pairs. However, the absence of *Wolbachia* in our investigation can also be the result of using somatic tissue, instead of reproductive tissue, as the primary source of DNA. Use of somatic tissue may decrease detection of *Wolbachia*. As a result, *Urbanus bernikerni* could be infected, but the somatic tissue tested could incorrectly test as “absent” for all 13 specimens examined, though we think that this is unlikely. To date,

somatic infections have been reported in species including some in the genus *Drosophila* [58] and have been proposed to increase the likelihood of horizontal transmission routes [59]. There have been reported cases of infection being isolated to reproductive tissues in *Glossina morsitans* (Diptera, Glossinidae) females [58]. To further investigate the potential role of *Wolbachia* and speciation within the *Urbanus bernikerni* complex, reproductive, as well as somatic tissues need to be analyzed.

Habitat distribution of the *Urbanus bernikerni* complex in the ACG

Now that we recognize three species in *Urbanus bernikerni* complex, it begins to be possible to garner some ecological sense from their intertwined and overlapping distribution within ACG. ACG is divided roughly south-north by the cloud forest-topped Cordillera Guanacaste, with Pacific coastal dry forest on the western side, Caribbean rain forest on the eastern side, and a variety of intergrades between these three major parapatric ecosystems (Figure 2). *Urbanus viterboana* is restricted to the upper cooler slopes of the volcanos, and is parapatric with the three *Urbanus bernikerni* complex species restricted to the lowlands from about 600–800 m down to sea level. *Urbanus ehakernae* is clearly a rain forest denizen and seems to have remained there (Figure 2). While *Urbanus segnestami* (Figure 2) is unambiguously able to tolerate the climatic conditions of Pacific dry forest, its omnipresence throughout ACG middle to low elevations may be a recent artifact of logging, clearing, cultivating, and pasturing of the rain forest side of ACG, an anthropomorphic force that severely insulates, heats, and dries the previously much shadier, cooler, and wetter Caribbean ground-level rain forest side of ACG. These open agroscape habitats are rich in herbaceous ruderal Asteraceae, plants fed on by the caterpillars of all three members of the *Urbanus bernikerni* complex. This presumed invasion of the (largely destroyed) rain forest ecosystem at ground level by ACG dry forest species has been repeated by many other species of insects and plants. *Urbanus bernikerni* itself (Figure 2) also appears to have originally occupied the moister and cooler marginal/intergrade portions of the ACG dry forest with the rain forest nudging in from the east, or local more moist/evergreen parts of dry forest. It will be literally many centuries before the ACG forest-in-restoration once again mimics an intact old growth forest, with its very different physical characteristics from what we see today; and we can then document what would be the “true” ecological distribution of these three sibling species in ACG. That is, however, if the habitat island of wildland ACG as a whole is not thoroughly swamped by the population and community dynamics of these three species and their host Asteraceae in the ocean of agroscape occupying Central America.

Conclusion

Making use of DNA barcode libraries to generate hypotheses for provisional boundaries of unrecognized species can allow researchers to allocate their exploratory resources more efficiently. The discovery of the *Urbanus bernikerni* complex demonstrates that using an integrative framework, led by genetic analyses, can provide adequate evidence to confirm the presence of as-yet undescribed cryptic species. Having prior knowledge of habitat distribution played a key role in identifying this complex, and such data are highly recommended for future work. We are also careful to note that results such as these render casual specimen identification yet more difficult; we do not, for example, know which, if any, of these three cryptic species is the same as what is generally known as *Urbanus belli* throughout the Neotropics, since the type specimen for that name has not been DNA barcoded. Thus, extending this study to include closely related congeneric species over a broader geographical distribution is of primary interest.

Availability of supporting data

The data sets supporting the results of this article are available in the Dryad repository, doi:10.5061/dryad.pj561 and <http://datadryad.org/review?wfid=29486&token=9cf99242-b5e5-4baa-bac4-223762f92cdb> [60].

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

CB, DHJ, WH, JMB and MH conceived the idea and designed experiments. CB, DHJ, WH, JMB, SS and MH performed experiments. CB, DHJ, WH, JMB, JFG, SS and MH analyzed and interpreted data. All authors wrote and edited the manuscript. All authors read and approved the final manuscript.

Acknowledgments

We gratefully acknowledge the team of ACG parataxonomists [13,60] who found and reared the specimens used in this study, and the team of biodiversity managers who keep the ACG forests that host these butterflies alive. We thank Karie Darrow for preparing Figure 3 and Donald Harvey for dissecting many genitalia. We also thank Teresa Crease and Alex Smith for providing advice and editorial assistance in the early stages of this research. The study has been supported by U.S. National Science Foundation grants BSR 9024770 and DEB 9306296, 9400829, 9705072, 0072730, 0515699, and grants from the Wege Foundation, International Conservation Fund of Canada, Jessie B. Cox Charitable Trust, Blue Moon Fund, Guanacaste Dry Forest Conservation Fund, Permian Global, Area de Conservación Guanacaste, and University of Pennsylvania (DHJ). This project was funded by the Government of Canada through Genome Canada and the Ontario Genomics Institute (OGI-050) through the Biomonitoring 2.0 project (www.biomonitoring2.org) (MH) and by a Discovery Grant from Natural Sciences and Engineering Research Council of Canada (MH).

Author details

¹Biodiversity Institute of Ontario & Department of Integrative Biology, University of Guelph, Guelph, Ontario N1G 2W1, Canada. ²Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ³Department of Entomology, National Museum of Natural History, Smithsonian Institution, P.O.Box 37012, Washington, DC 20013-7012, USA. ⁴Department of Microbiology, Mansoura University, 35516 Mansoura, Egypt.

Received: 19 February 2014 Accepted: 17 June 2014

Published: 9 July 2014

References

1. Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, Ingram KK, Das I: **Cryptic species as a window on diversity and conservation.** *Trends Ecol Evol* 2007, **22**:148–155.
2. Bidochka MJ, Kamp AM, Lavender TM, Dekoning J, DeCraets JNA: **Habitat association in two genetic groups of the insect-pathogenic fungus *Metarhizium anisopliae*: uncovering cryptic species?** *Appl Environ Microbiol* 2001, **67**:1335–1342.
3. Geller JB: **Decline of a native mussel masked by sibling species invasion.** *Conserv Biol* 1999, **13**:661–664.
4. Smith MA, Wood DM, Janzen DH, Hallwachs W, Hebert PDN: **DNA barcodes affirm that 16 species of apparently generalist tropical parasitoid flies (Diptera, Tachinidae) are not all generalists.** *Proc Natl Acad Sci U S A* 2007, **104**:4967–4972.
5. Tan DSH, Ang Y, Lim GS, Ismail MRB, Meier R: **From 'cryptic species' to integrative taxonomy: an iterative process involving DNA sequences, morphology, and behaviour leads to the resurrection of *Sepsis pyrrhosoma* (Sepsidae: Diptera).** *Zool Scr* 2010, **39**:51–61.
6. Janzen DH, Hallwachs W, Harvey DJ, Darrow K, Rougerie R, Hajibabaei M, Smith MA, Bertrand C, Gamboa IC, Espinoza B, Sullivan JB, Decaens T, Herbin D, Chavarria LF, Franco R, Cambronero H, Rios S, Quesada F, Pereira G, Vargas J, Guadamuz A, Espinoza R, Hernandez J, Rios L, Cantillano E, Moraga R, Moraga C, Rios P, Rios M, Calero R, et al: **What happens to the traditional taxonomy when a well-known tropical saturniid moth fauna is DNA barcoded?** *Invert Syst* 2012, **26**:478–505.
7. Hebert PDN, Ratnasingham S, DeWaard JR: **Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species.** *Proc R Soc Lond B Biol Sci* 2003, **270**:S96–S99.
8. Janzen DH, Hallwachs W, Blandin P, Burns JM, Cadiou JM, Chacon I, Dapkey T, Deans AR, Epstein ME, Espinoza B, Franclemont JG, Haber WA, Hajibabaei M, Hall JPW, Hebert PDN, Gauld ID, Harvey DJ, Hausmann A, Kitching JJ, Lafontaine D, Landry JF, Lemaire C, Miller JY, Miller JS, Miller L, Miller SE, Montero J, Munroe E, Green SR, Ratnasingham S, et al: **Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity.** *Mol Ecol Res* 2009, **9**:1–26.
9. Burns JM, Janzen DH, Hajibabaei M, Hallwachs W, Hebert PDN: **DNA barcodes of closely related (but morphologically and ecologically distinct) species of skipper butterflies (Hesperiidae) can differ by only one to three nucleotides.** *J Lepid Soc* 2007, **61**:138–153.
10. Burns JM, Janzen DH, Hajibabaei M, Hallwachs W, Hebert PDN: **DNA barcodes and cryptic species of skipper butterflies in the genus *Perichares* in Área de Conservación Guanacaste, Costa Rica.** *Proc Natl Acad Sci U S A* 2008, **105**:6350–6355.
11. Janzen DH, Hallwachs W, Burns JM, Hajibabaei M, Bertrand C, Hebert PDN: **Reading the complex skipper butterfly fauna of one tropical place.** *PLoS ONE* 2011, **6**(8):e19874.
12. Grishin NV, Janzen DH, Hallwachs W: **Hiding behind gaudy looks, a new Central American species of *Phareas* (Hesperiidae: Eudaminae).** *J Lepid Soc* 2013, **67**:161–174.
13. Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W: **Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptus fulgurator*.** *Proc Natl Acad Sci U S A* 2004, **101**:14812–14817.
14. Onyabe DY, Conn JE: **Intragenomic heterogeneity of a ribosomal DNA spacer (ITS2) varies regionally in the neotropical malaria vector *Anopheles nuneztovari* (Diptera: Culicidae).** *Insect Mol Biol* 1999, **8**:435–442.
15. Song J, Shi L, Li D, Sun Y, Niu Y, Chen Z, Luo H, Pang X, Sun Z, Liu C, Lv A, Deng Y, Larson-Rabin Z, Wilkinson M, Chen S: **Extensive pyrosequencing reveals frequent intragenomic variations of internal transcribed spacer regions of nuclear ribosomal DNA.** *PLoS ONE* 2012, **7**(8):e43971.
16. Müller T, Philippi N, Dandekar T, Schultz J, Wolf M: **Distinguishing species.** *RNA* 2007, **13**:1469–1472.
17. Wolf M, Chen S, Song J, Ankenbrand M, Müller T: **Compensatory base changes in ITS2 secondary structures correlate with the biological species concept despite intragenomic variability in ITS2 sequences – a proof of concept.** *PLoS ONE* 2013, **8**(6):e66726.
18. Baldo L, Hotopp JCD, Jolley KA, Bordenstein SR, Biber SA, Choudhury RR, Hayashi C, Maiden MCJ, Tettelin H, Werren JH: **Multilocus sequence typing**

- system for the endosymbiont *Wolbachia pipientis*. *Appl Environ Microbiol* 2006, **72**:7098–7110.
19. Bordenstein SR: **Symbiosis and the Origin of Species**. In *Insect Symbiosis*. Edited by Bourtzis K, Miller T. New York: CRC Press; 2003:283–304.
 20. Werren JH, Baldo L, Clark ME: *Wolbachia*: master manipulators of invertebrate biology. *Nat Rev Microbiol* 2008, **6**:741–751.
 21. Janzen DH, Hallwachs W: *Area de Conservacion Guanacaste (ACG), Northwestern Costa Rica, Caterpillars, Pupae, Butterflies & Moths*. 2013 [http://janzen.bio.upenn.edu/caterpillars/database.lasso]
 22. Ratnasingham S, Hebert PDN: **BOLD: The Barcode of Life Data System**. *Mol Ecol Notes* 2007, **7**:355–364 (www.barcodinglife.org).
 23. Cho SW, Mitchell A, Regier JC, Mitter C, Poole RW, Friedlander TP, Zhao SW: **A highly conserved nuclear gene for low-level phylogenetics *Elongation factor 1 alpha* recovers morphology-based tree for Heliothine moths**. *Mol Biol Evol* 1995, **12**:650–656.
 24. Nazari V, Zakharov EV, Sperling FAH: **Phylogeny, historical biogeography, and taxonomic ranking of Parnassiinae (Lepidoptera, Papilionidae) based on morphology and seven genes**. *Mol Phylogenet Evol* 2007, **42**:131–156.
 25. White TJ, Bruns T, Lee S, Taylor JW: **Amplification and Direct Sequencing of Fungal Ribosomal RNA Genes for Phylogenetics**. In *PCR Protocols: A Guide to Methods and Applications*. Edited by Innis MA, Gelfand DH, Sninsky JJ, White TJ. New York: Academic Press, Inc; 1990:315–322.
 26. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace JM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG: **ClustalW and ClustalX version 2**. *Bioinformatics* 2007, **23**:2947–2948.
 27. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0**. *Mol Biol Evol* 2007, **24**:1596–1599.
 28. Nylander JAA: *MrModeltest v2. Program Distributed by the Author*. Evolutionary Biology Centre, Uppsala University; 2004.
 29. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models**. *Bioinformatics* 2003, **19**:1572–1574.
 30. Ewing B, Green P: **Base-calling of automated sequencer traces using phred. II. Error probabilities**. *Genome Res* 1998, **8**:186–194.
 31. Ewing B, Hillier L, Wendl MC, Green P: **Base-calling of automated sequencer traces using phred. I. Accuracy assessment**. *Genome Res* 1998, **8**:175–185.
 32. Schmieder R, Edwards R: **Quality control and preprocessing of metagenomic datasets**. *Bioinformatics* 2011, **27**:863–864.
 33. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R: **UCHIME improves sensitivity and speed of chimera detection**. *Bioinformatics* 2011, **27**:2194–2200.
 34. Li W, Godzik A: **Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences**. *Bioinformatics* 2006, **22**:1658–1659.
 35. Bray JR, Curtis JT: **An ordination of the upland forest communities of southern Wisconsin**. *Ecol Monogr* 1957, **27**:326–349.
 36. Clarke KR, Gorley RN: *PRIMER v6: User Manual/Tutorial*. Plymouth: PRIMER-E; 2006.
 37. Swofford DL: *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4*. Sunderland, MA: Sinauer Associates; 2002.
 38. Gutell RR, Larsen N, Woese CR: **Lessons from an evolving ribosomal-RNA-16S and 23S ribosomal-RNA structures from a comparative perspective**. *Microbiol Rev* 1994, **58**:10–26.
 39. Seibel PN, Muller T, Dandekar T, Wolf M: **Synchronous visual analysis and editing of RNA sequence and secondary structure alignments using 4SALE**. *BMC Res Notes* 2008, **1**:91.
 40. Hayward KJ: **Los Pyrginae argentinos. Adiciones y anotaciones (Lep. Grypocera)**. *An Soc Cient Argentina* 1935, **119**(6):256–261.
 41. Evans WH: *A catalogue of the American Hesperidae indicating the classification and nomenclature adopted in the British Museum (Natural History), Part II, Pyrginae, Section I*. London: British Museum; 1952.
 42. Ehrmann GA: **New tropical American Hesperidae**. *Can Entomol* 1907, **39**:317–323.
 43. Steinhauser SR: **A revision of the *proteus* group of the genus *Urbanus* Hubner, Lepidoptera: Hesperidae**. *Bull Allyn Mus* 1981, **62**:1–41.
 44. Schultz J, Maisel S, Gerlach D, Muller T, Wolf M: **A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota**. *RNA* 2005, **11**:361–364.
 45. Behere GT, Tay WT, Russell JA, Heckel DG, Appleton BR, Kranthi KR, Batterham P: **Mitochondrial DNA analysis of field populations of *Helicoverpa armigera* (Lepidoptera: Noctuidae) and of its relationship to *H. zea***. *BMC Evol Biol* 2007, **7**:117.
 46. Russell JA, Goldman-Huertas B, Moreau CS, Baldo L, Stahlhut JK, Werren JH, Pierce NE: **Specialization and geographic isolation among *Wolbachia* symbionts from ants and lycaenid butterflies**. *Evolution* 2009, **63**:624–640.
 47. Yao H, Song J, Liu C, Luo K, Han J, Li Y, Pang X, Xu H, Zhu Y, Xiao P, Chen S: **Use of ITS2 region as the universal DNA barcode for plants and animals**. *PLoS ONE* 2010, **5**(10):e13102.
 48. Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W: **Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for *Fungi***. *PNAS* 2012, **109**(16):6241–6246.
 49. Alvarez I, Wendel JF: **Ribosomal ITS sequences and plant phylogenetic inference**. *Mol Phylogenet Evol* 2003, **29**:417–434.
 50. Arnheim N, Krystal M, Schmickel R, Wilson G, Ryder O, Zimmer E: **Molecular evidence for genetic exchanges among ribosomal genes on non-homologous chromosomes in man and ape**. *Proc Natl Acad Sci U S A* 1980, **77**:7323–7327.
 51. Hershkovitz MA, Zimmer EA, Hahn WJ: **Ribosomal DNA Sequences and Angiosperm Systematics**. In *Molecular Systematics and Plant Evolution*. Edited by Hollingsworth PM, Bateman RM, Gornall RJ. London: Taylor & Francis Limited; 1999:268–326.
 52. Smith MA, Woodley NE, Janzen DH, Hallwachs W, Hebert PDN: **DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera : Tachinidae)**. *Proc Natl Acad Sci U S A* 2006, **103**:3657–3662.
 53. Fisher BL, Smith MA: **A revision of Malagasy species of *Anochetus* Mayr and *Odontomachus* Latreille (Hymenoptera: Formicidae)**. *PLoS ONE* 2008, **3**(5):e1787.
 54. Mayol M, Rossello JA: **Why nuclear ribosomal DNA spacers (ITS) tell different stories in *Quercus***. *Mol Phylogenet Evol* 2001, **19**:167–176.
 55. Kircher M, Kelso J: **High-throughput DNA sequencing - concepts and limitations**. *Bioessays* 2010, **32**:524–536.
 56. Coleman AW: **ITS2 is a double-edged tool for eukaryote evolutionary comparisons**. *Trends Genet* 2003, **19**:370–375.
 57. Coleman AW: **Is there a molecular key to the level of “biological species” in eukaryotes? A DNA guide**. *Mol Phylogenet Evol* 2009, **50**:197–203.
 58. Dobson SL, Bourtzis K, Braig HR, Jones BF, Zhou W, Rousset F, O'Neill SL: ***Wolbachia* infections are distributed throughout insect somatic and germ line tissues**. *Insect Biochem Mol Biol* 1999, **29**(2):153–160.
 59. Werren JH, Windsor D, Guo L: **Distribution of *Wolbachia* among neotropical arthropods**. *Proc Biol Sci* 1995, **262**:197–204.
 60. Bertrand C, Janzen DH, Hallwachs W, Burns JM, Gibson JF, Shokralla S, Hajibabaei M: **Data from: Mitochondrial and nuclear phylogenetic analysis with Sanger and next-generation sequencing shows that, in Area de Conservacion Guanacaste, northwestern Costa Rica, the skipper butterfly named *Urbanus belli* (family Hesperidae) comprises three morphologically cryptic species**. *Dryad Digital Repository* doi:10.5061/dryad.pj561.

doi:10.1186/1471-2148-14-153

Cite this article as: Bertrand et al.: Mitochondrial and nuclear phylogenetic analysis with Sanger and next-generation sequencing shows that, in Area de Conservación Guanacaste, northwestern Costa Rica, the skipper butterfly named *Urbanus belli* (family Hesperidae) comprises three morphologically cryptic species. *BMC Evolutionary Biology* 2014 **14**:153.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

