PHYLOGENETICS AND INTROGRESSION OF *HABRONATTUS* JUMPING SPIDERS USING TRANSCRIPTOMES (*ARANEAE: SALTICIDAE*)

by

Geneviève Leduc-Robert

B.Sc., The University of British Columbia, 2011

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate and Postdoctoral Studies

(Zoology)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

December 2015

Abstract

Habronattus is a diverse genus of jumping spiders with complex courtship displays and colourful ornaments in males. A well-resolved species phylogeny would provide an important framework to study these traits, but has not yet been achieved because of conflicting signals from the few genes available. While such discordant gene trees could be the result of deep coalescence in the recently diverged group, there are many indications that hybridization may have occurred and could be the source of conflict. To infer *Habronattus* phylogenetic relationships and to investigate the cause of gene tree discordance, we assembled transcriptomes for 34 *Habronattus* species and 2 outgroups. We conducted a concatenated phylogenetic analysis using Maximum Likelihood for 2.41 Mb of nuclear data and for 12.33 kb of mitochondrial data. The concatenated nuclear phylogeny was resolved with high bootstrap support (95-100%) at most nodes with some uncertainty surrounding the relationships of *H. icenoglei*, *H. cambridgei,* and *H. oregonensis,* and *Pellenes* cf. *levii*. There are several nodes of the mitochondrial phylogeny that are incongruent to the nuclear phylogeny and indicate possible mitochondrial introgression: the internal relationships of the *americanus* and the *coecatus* group, the relationship between the *altanus, decorus, banksi*, and *americanus* group, and between *H. clypeatus* and the *coecatus* group. To determine the extent of incomplete lineage sorting (ILS) and introgression, we analyzed gene tree discordance for loci longer than 1 kb using Bayesian Concordance Analysis (BCA) for the *americanus* group (679 loci) and the *viridipes*/*clypeatus*/*coecatus* (VCC) group (517 loci) and found high levels of genetic discordance, especially in VCC group. Finally, we tested specifically for nuclear introgression in the concatenated nuclear matrix with Patterson's D statistics and D_{FOIL}. We found nuclear introgression resulting in substantial admixture between *americanus* group species, and between *H.* sp. (ROBRT) and the *clypeatus* group, and more

minimal nuclear introgression between the *clypeatus* group and the *coecatus* group, and between the *americanus* group and several distant species. Our results indicate that hybridization may have been historically common between phylogenetically distant species of *Habronattus*, and that reproductive isolation is yet to be complete across the *Habronattus* phylogeny.

Preface

The collection of specimens was conducted as a team led by W.P. Maddison in Arizona and Mexico. Other specimens were already collected by David Maddison, Junxia Zhang, Damian Elias, and Marshal Hedin. I designed the study with the help of W.P. Maddison. I completed all the lab work and data analysis, except for the Bioanalyzer step completed by Joanne Denny, the cDNA library preparation completed by Debbie Adams and Anastasia Kuzmin and the Illumina sequencing step competed by Anastasia Kuzmin. There were steps in alignment and sequence processing completed with custom modules, designed by W.P. Maddison, in the Mesquite Gataga package. I prepared the manuscript with input from W.P. Maddison.

Table of contents

List of tables

List of figures

Acknowledgements

I'd like to thank Wayne Maddison for his insights, encouragement, and contagious enthusiasm. His expertise on *Habronattus* and phylogenetics were instrumental to this thesis. Beyond his help with this project, I look to him as a role model of what a scientist, mentor, and educator can be. Thank you for setting such a strong example.

A special thanks to Heather Proctor, Sam Evans, and Abraham Meza for help collecting specimens for this project in August 2013, and to Marshal Hedin and Megan Porter, who contributed sequencing data.

Thank you to Saemundur Sveinsson for his patient help as I started navigating bioinformatics and sequencing data for the first time. Thanks to Allyson MisCampbell and Carol Ritland from the Forestry Genomic Data Centre at UBC, and Jeff Richards and his students, for their help and the resources they shared in the lab. I am also grateful to Anastasia Kuzmin for being so helpful and collaborative throughout the sequencing process. Thanks to Westgrid (CanadaComputes) for access to their computer clusters.

Thank you to my committee Sally Otto, Loren Rieseberg, and Darren Irwin, and to Maddison lab members past and present, Edy Piascik, Junxia Zhang, Gwylim Blackburn, and Sam Evans, for their helpful feedback and ideas.

Thank you to Sean Naman for his daily positive presence, kindness, and inspiration and to the friends who have made these years so enjoyable. I'm especially thankful to my parents, Josée Leduc and André Robert, for their encouragement and support, and for instilling in me a love of learning and of the natural world.

This work was funded by an NSERC grant to W.P. Maddison,

1. Introduction

A major challenge in phylogenetics is accurately inferring species relationships in recently and rapidly diverged clades (Degnan and Rosenberg 2009). Because of recent divergence, these clades are prone to incomplete lineage sorting (ILS) and are likely to have weak reproductive barriers that are susceptible to hybridization (Donnelly et al. 2004; Grant et al. 2005), two processes that contribute to signals discordant to the species tree. A clade's true evolutionary history is therefore reflected both in the species tree and in discordant signals (Maddison 1997; Baum 2007). Determining the relative contributions of introgression and ILS to genetic discordance can help us understand the relationship between hybridization and divergence during diversification, a key field of research in evolution (Mallet 2005; Abbott et al. 2013; Seehausen 2004; Baack and Rieseberg 2007). Hybridization can have several different effects: it can accelerate speciation through reinforcement (Turelli et al. 2001), reduce genetic divergence between hybridizing species (Anderson 1949), or it can introduce new genetic variation at a faster rate than would be possible by mutation alone, potentially sharing adaptive loci across lineages and facilitating diversification (Slatkin 1985; Twyford and Ennos 2012; Rieseberg et al. 2003). While adaptive introgression and hybrid speciation are documented as important and common events in plants, it is only recently that examples of adaptive introgression have been described in animals (Seehausen 2004; Mallet 2005). However, it is still unclear if introgression plays a major or minor role in the diversification of animals. A better characterization of nuclear introgression in different clades is needed to determine the relative contribution of introgression to adaptive genetic variation.

Because discordant gene trees produced by introgression or ILS are difficult to distinguish (Joly et al. 2009), a large sample of nuclear genes is required to identify nuclear

introgression. Until recently, phylogenetic datasets have included only a small sample of genes; thus, nuclear introgression is poorly described in most organisms relative to mitochondrial introgression, which may leave a different historical signature that does not correspond to nuclear introgression (Toews and Brelsford 2012). This issue can be addressed using phylogenomic datasets, which can successfully resolve species trees with high levels of discordance (Dunn et al. 2008; Lemmon and Lemmon 2012), and which can also identify patterns of historical introgression (Yu et al. 2011; Kronforst et al. 2006) because characteristics unique to introgression emerge in genome-wide data. For example, two gene trees discordant to the species tree (in a rooted four-taxon tree) are predicted to occur at equal frequencies when ILS is the only cause of discordance (Durand et al. 2011). Because ILS is a random process, it should not favour one particular discordant topology or another. This is not the case if introgression has occurred, because the discordant gene trees which link together species that introgressed as sister taxa should be more frequent due to a higher frequency of shared (via introgression) loci between those two species. Therefore, asymmetric frequencies of discordant signals favouring close relationships between potentially hybridizing species can be used as a diagnosis for introgression with a large enough set of gene trees (Baum 2007; Larget et al. 2010) or SNPs (Durand et al. 2011; Martin et al. 2015; Pease and Hahn 2015). Coalescence times can also differentiate introgression from ILS (Joly et al. 2009). Gene trees produced by ILS should coalesce prior to the time of speciation while introgression should result in coalescence following speciation (Joly et al. 2009).

Habronattus is a young jumping spider (*Salticidae*) genus comprised of about 100 described species that diversified primarily in North America, with some species present as far south as Costa Rica. They possess an impressive diversity of colourful male ornaments and

courtship behaviours, amongst the most complex found in arthropods (Peckham and Peckham 1889; Peckham and Peckham 1890; Griswold 1987; Maddison and McMahon 2000; Maddison and Hedin 2003; Elias et al. 2012). Multimodal courtship displays can involve up to 23 male ornaments, motion displays, and vibratory signals, arranged temporally into motifs (Elias et al. 2012). As a result of this complex behaviour, *Habronattus* is emerging as a model to study the role of sexual selection in divergence (Hebets and Maddison 2005; Scheidemantel 1997; Elias 2006; Blackburn and Maddison 2014; Maddison and McMahon 2000; Masta and Maddison 2002), the evolution of sex chromosomes (Maddison and Leduc-Robert 2013; Maddison 1982), and arachnid visual systems (Zurek et al. 2015).

Several clades within *Habronattus* are morphologically distinctive and have been partially phylogenetically resolved using Ef1-a (nuclear gene) and 16SND1 (mitochondrial ribosomal subunit and adjacent ND1 gene) (Maddison and Hedin 2003) and morphological traits (Griswold 1987). The genus as a whole is defined by an elbowed tegular apophysis (part of the male palp), which is secondarily lost in the *coecatus* group (Maddison and Hedin 2003; Griswold 1987), and shorter first legs in most species compared to sister genus *Pellenes*. The *americanus*, *agilis*, *amicus*, *tranquillus* (together referred to as the AAT clade), and *dorotheae* groups were first inferred based on morphology (Griswold 1987) and later confirmed with molecular data (Maddison and Hedin 2003). The *decorus*, *coecatus*, and *viridipes* groups began to take form in the morphological phylogeny (Griswold 1987) but only fully emerged with molecular data (Maddison and Hedin 2003). In particular, the *viridipes* group was further divided into the *clypeatus* group (primarily southwestern species) and the *viridipes* group (primarily northern species) based on molecular data (Maddison and Hedin 2003). Despite this success, other aspects of *Habronattus* phylogeny have been difficult to resolve because of insufficient data and

widespread discordance (Maddison & Hedin 2003; Griswold 1987). The internal relationships of most species groups and the deeper phylogenetic structure of the genus remain poorly resolved. As a result, the analysis of character traits, such as male ornaments and courtship behaviour, within a phylogenetic framework has been limited.

There are many indications that introgression may be partially responsible for poor phylogenetic resolution due to gene tree discordance. There have been several molecular hints of mitochondrial introgression, both between distant species (Maddison and Hedin 2003), and between sympatric and closely related species that form mitochondrial clades based on geographic proximity rather than by morphospecies (Hedin and Lowder 2009; Maddison and Hedin, unpubl.). Introgression could also explain why some morphologically divergent species group together with strong support in the nuclear phylogeny (Maddison and Hedin 2003). The possibility of introgression is further supported by behavioural tests, which found that females from different populations of *H. pugillis* can have preferences for foreign males with divergent courtship displays, a possible result of antagonistic coevolution between the sexes (Hebets and Maddison 2005). If widespread hybridization is possible, in particular between more distant species that are expected to be reproductively isolated, traits novel to the recipient lineages could potentially introgress. For example, sexually selected ornaments in males from divergent and allopatric *H. pugillis* populations exhibit convergence that could be explained by introgression (Maddison and McMahon 2000). To determine if historical hybridization in *Habronattus* has been significant and if its evolutionary effects have been creative, neutral, or destructive, the extent of hybridization across the group and the amount of nuclear introgression resulting from this hybridization first needs to be quantified.

We set out to further resolve the *Habronattus* phylogeny and to determine the extent of introgression in the clade. We collected transcriptome data for 34 *Habronattus* species and two outgroups, the first genomic dataset assembled for *Salticidae* spiders. With a well-resolved phylogenetic tree, we were then able to use a phylogenetic approach to determine the extent of nuclear and mitochondrial introgression in the group. We focused on the *americanus* group and the VCC group, two of the most diverse species groups within *Habronattus*. To investigate nuclear introgression, we first conducted a Bayesian Concordance Analysis (BCA) to investigate discordance (caused by either ILS or introgression) in gene trees, and we then applied Patterson's D statistic and D_{FOIL} tests to explicitly test allele patterns for introgression. We resolved most nodes of the phylogeny with high support and identified several instances of hybridization in the group.

2. Methods

2.1. Taxon sampling

We sampled a total of 36 species, including representatives from all major clades within *Habronattus,* and 2 outgroups (Appendix 1). We sampled more deeply in species-rich groups (i.e., *viridipes/clypeatus/coecatus* group, and the *americanus* group), and prioritized the *coecatus* group because of possible introgression (Maddison and Hedin 2003). The outgroup *Pellenes* cf. *levii* is from the genus sister to *Habronattus* (subtribe *Harmochirinae*), and *Evarcha proszjinski* is more distantly related (subtribe *Plexippinae*). Both the genera *Evarcha* and *Pellenes* have been previously used in phylogenies to support *Habronattus* as monophyletic (Maddison and Hedin 2003). The undescribed species *H*. sp. (ROBRT) and *H*. sp. (CHMLA) retain their 5 letter coded names from Maddison and Hedin (2003) and *H*. cf. *dossenus* (Silvercity) of Maddison and Hedin (2003) is renamed here as undescribed species *H*. sp. (SLCTY).

This is the first phylogenetic analysis to include the undescribed species *H*. sp. (ESTU), *H*. sp. (BLNDI), and *H*. sp. (SUNGL). Because these species have yet to be fully described, they are given new 4-5 coded letter names until a complete species description is made in the future. *H*. sp. (ESTU) was found near water at low tide along the estuaries of Puerto Peñasco, Mexico and has an *americanus* group style courtship behaviour. *H*. sp. (BLNDI), which was caught near *H*. sp. (ESTU) but in a drier habitat, resembles a blond *H. pyrrithrix* (a species also found sympatrically) with a similar distinctive bright red face. *H*. sp. (SUNGL) was found in high desert grass clumps near Mount Hopkins, Arizona. It resembles a paler and heavier bodied *H. pugillis,* and has a bright orange and blue face.

Specimens were collected from 2012 to 2014 in various locations in Canada (British Columbia, Ontario, and Alberta), the USA (Arizona, California, and New Mexico), Mexico

(Sonora and Jalisco), and Panama (localities of all specimens are listed in Appendix 1). *Habronattus* are primarily ground-dwellers found on rocks, leaf litter or grasses; therefore, we collected specimens by catching them off the ground and we used a beat sheet method for the few species that were arboreal. Each specimen was photographed alive in the field. Adult male specimens were preferred because they are easier to distinguish morphologically between species. We resorted to adult females for *H.* sp. (ESTU), *H.* sp. (SUNGL), *H. captiosus,* and *H.* sp. (SLCTY) because males were not available. Both a male and a female specimen were included for *H. ophrys* and *H. festus* in an effort to assemble a more complete reference transcriptome.

H. paratus DNA was preserved in 95% EtOH. All other specimens were killed by submersion in RNAlater for RNA preservation. To maximize tissue exposure, the cephalothorax and abdomen were opened immediately upon submersion. All specimens were stored at -20°C. Legs and the male palps or the female epigynum were preserved separately as vouchers (stored at the Beaty Biodiversity Museum at the University of British Columbia).

2.2. Molecular extractions and sequencing

We chose RNA-seq because the assembled transcripts are long enough to permit a gene tree approach that is difficult to implement with other common phylogenomic methods (e.g., RADseq, targeted sequence capture). Transcriptomes also avoid uninformative sequences that may form a major part of the large *Habronattus* genome (haploid genome size = 5.73 pg, Gregory and Shorthouse 2003) and that would be assembled using whole-genome methods. Total RNA was extracted from whole specimens using a combination of TRIzol extraction (Life Technologies) and RNeasy Mini Kit (Qiagen) for RNA purification and DNAse digestion. DNA

was extracted from the legs and abdomen of *H. paratus* using a QIAamp DNA Investigator Kit (Qiagen). Samples were selected for sequencing based on three criteria: (1) quantity (minimum 2 ug) as measured by a spectrophotometer, (2) purity, (minimum 260/280 = 1.8 for RNA and 260/280 = 2 for DNA, minimum 230/260 = 1.5) and (3) RNA integrity, assessed on a 2100 Bioanalyzer. RNA samples were kept at -80°C and DNA samples were kept at -20°C until the cDNA library preparation step.

Libraries were constructed with BIO-O NEXTflex Library Prep Kits (Bioo Scientific, Inc.) with insert sizes averaging 220bp for RNA and 300 bp for DNA, and sequenced as 100bp paired-end reads on an Illumina HiSeq 2000 (Illumina, Inc.) at the Biodiversity Research Centre Next Generation Sequencing Facilities (University of British Columbia). We aimed for at least 20 million paired reads per species, with the exception of potential reference transcriptomes *H. ophrys* and *H. festus,* which were more deeply sequenced (approximately half a sequencing lane each, see Appendix 2 for sequencing summary).

H. signatus and *H. ustulatus* were prepared at UC San Diego (M. Hedin, unpubl.) and sequenced as 50 bp paired end sequences, and *H.* sp. (SLCTY) was prepared at the University of South Dakota (M. Porter, unpubl.) and sequenced as 150 bp paired end sequences.

2.3 Sequence read filtering and trimming

Any sequence read with an average Phred score under Q=30 was discarded. All remaining reads were quality checked with FASTQC V0.10.1 (Andrews 2010).
Terminal nucleotides were trimmed using fastq-mcf from ea-utils (Aronesty 2011) if they had a score below Q=30 or if they were sequencing adaptors. Reads that were 95% or more homopolymer were discarded and any suspected contaminants, detected from the GC content

curve of FASTQC, were trimmed using PRINSEQ-lite (Schmieder and Edwards 2011). Any read shorter than 33 bp after trimming was discarded. Contamination was ruled out as the source of overrepresented sequences when the top blastn hits returned for these sequences were *Habronattus* mitochondrial and ribosomal RNA sequences.

2.4 Reference transcriptome

Reference transcriptomes were assembled for *H. festus* and *H. ophrys* in Trinity RNA-seq_v20140717 (Haas et al. 2013) with the command:
Trinity.pl --seqType fq --left leftreads.fq –right rightreads.fq --JM 110G --CPU 12 --inchworm_cpu 12 --bflyCPU 12 --min_contig_length 200 –-kmer_cov 2

Both assemblies were similar in size and quality, so we selected *H. ophrys* as the reference transcriptome because of predicted ease of obtaining this species for future studies.

We filtered transcripts from the *H. ophrys* prior to using the transcriptome as a reference for subsequent assemblies according to the steps outlined in Figure 1. We determined transcript abundance with RSEM v1.2.19 (Li and Dewey 2011) and kept only the most abundant transcript variant per gene. Any remaining redundant transcripts identified with CD-HIT-EST (Fu et al. 2012) with a similarity threshold of 95% were removed. To decrease the likelihood of paralogous genes rather than true orthologs assembling on a reference transcript during reference-based assemblies, we conducted an all-versus-all BLAST with all remaining *H. ophrys* contigs and removed any contig with a contig other than itself as a significant hit (blastx, evalue = 10^{-3}). To set codon positions, the reference transcriptome was scanned for open reading frames (ORFs) using TransDecoder_r20131110 (Haas et al. 2013) and the longest ORF of a transcript

was chosen as its protein coding region. If multiple non-overlapping coding regions were found on a transcript, those transcripts were split between coding regions.

We conducted a BLAST search of the entire *H. ophrys* transcriptome (evalue= 10^{-3}, min. HSP length = 33, max_target_seqs=20) to proteins from the SWISS-PROT database (Boeckmann et al. 2003), the African social velvet spider (*Stegodyphus mimosarum)* (Sanggaard et al. 2014) and the Brazilian white-knee tarantula (*Acanthoscurria geniculata*) (Sanggaard et al. 2014) to assess the quality and completeness of the transcriptome. The transcriptome was also BLASTed to the *H. oregonensis* mitochondrial genome (Masta and Boore 2004) and any significant hits were removed from the final reference nuclear transcriptome. The annotated *H. oregonensis* mitochondrial genome (Masta and Boore 2004), was used instead as the reference for all mitochondrial assemblies.

2.5 Reference-based assembly of transcriptomes

For each of the species other than *H. ophrys*, sequencing reads were first mapped to the *H. oregonensis* mitochondrial genome and the remaining unmapped reads were mapped to the *H. ophrys* reference transcriptome using CLC Genomics Workbench (CLC Bio) with assembly parameters: mismatch cost=2, insertion cost=3, deletion cost= 3, length fraction=0.5, similarity fraction=0.8. Polymorphisms were retained in the consensus sequence as ambiguous sites if the variant represented 30% or more of mapped alleles. Nuclear sequences with average coverage less than 5x were discarded, and contigs were split into fragments at any region where coverage was less than 5x. Only contigs longer than 200 bp following these steps were retained.

2.6 Sequence alignments

We converted species-based FASTA files into locus-based FASTA files and trimmed sequences of any remaining poly-A or poly-T tails using custom python scripts. If a loci was fragmented into multiple sequences for a species, and the lengths of those fragments when added together met our length cut-off, then those sequence fragments were scaffolded with "N"s to represent the missing data. Orthologous loci with a minimum average length of 200 bp (excluding "N"s) were aligned using MAFFT v7.058b (Katoh and Standley 2013) using L-INS-I and parameters --localpair --maxiterate 100. Nucleotides trailing from either end of an alignment (present in 30% or less of species) were trimmed and a sequence was discarded if it was 30% or less than the average sequence length for that alignment.

Matrices were partitioned as codon position 1,2,3, or non-coding based on Transdecoder results. To better understand the cause of gaps and ambiguous nucleotides in these alignments, we aligned a subset of transcripts with annotated sequences with known protein-coding regions matched using a BLAST search to the SWISS-PROT database, to Ef1-a (Maddison and Hedin 2003), and to the annotated mitochondrial genome (Masta and Boore 2004). All sites with gaps were caused by an insertion of a nucleotide in one or a few sequences. These insertions introduced gaps in the sequences that lacked the inserted nucleotide, causing highly unlikely frame-shifts in conserved genes. These insertions only occurred in the reference-based assemblies and never occurred in the reference sequence. Therefore, we inferred that these insertions of gaps and nucleotides were assembly errors rather than true insertions. As a result, columns with gaps were excluded if they were not present in the reference sequence whenever they were in a coding region (to avoid any frame-shift mutations) or if they were present in at least 50% of species in a non-coding region. If ambiguous sites constituted more than 3% of an

alignment, those alignments were excluded from phylogenetic analyses. High levels of ambiguous sites might not represent true heterozygosity, but rather multiple transcript variants or paralogs assembled on the same reference transcript. A 3% ambiguity threshold is very conservative and based on ambiguity levels in the sequences of three Ef1-a exons (exons described in Hedin and Maddison 2000). While exon 2 (ambiguity levels = 13%) and exon 1 (ambiguity levels =1.5%) have a top BLAST nr database hit to *Habronattus* ef1-a, ef1-a exon 3 (ambiguity levels = 21%) has more distant (not *Habronattus)* jumping spiders as its best BLAST nr database match. We therefore interpret high levels of ambiguity in a consensus sequence as a possible indication of multiple variants or paralogs contributing to the consensus sequence. All alignment trimming and partitioning steps were completed with the package Gataga (Maddison and Maddison, unpubl.) in Mesquite 3.02 (Maddison and Maddison 2015).

Because reference-based assemblies generate sequences aligned to the same reference transcripts, sequences assembled on the same reference are assumed to be orthologous. Filtering for paralogs in the *H. ophrys* transcriptome, reducing reference transcriptome redundancy, and removing sequences with high levels of ambiguous sites are expected to have sufficiently reduced the possibility of paralogs assembled on a reference sequence. To verify that paralogs were not an issue following these steps, gene trees were constructed for genes of lengths over 1 kb with a single search replicate in RAxML 7.7.9 (Stamatakis 2006). None of the trees produced had unusually long branches or phylogenetic structures obviously indicative of paralogy, so we considered these loci to be reliable orthologs.

2.7 Phylogenetic analyses

To determine optimal models of substitution, we analyzed in PartitionFinder v.1.1.1 (Lanfear et al. 2012) partitions for gene, codon positions (pos1, pos2, pos3) , and non-coding sites (N) for a sample of 20 nuclear genes and all mitochondrial genes using a greedy algorithm search and AIC model selection. All maximum likelihood phylogenetic searches were run with substitution model GTR+G+I, 20 search replicates, and 1000 bootstraps in RAxML 7.7.9 (Stamatakis 2006).

2.7.1. Nuclear and mitochondrial matrices

We conducted several phylogenetic analyses using different concatenated matrices, each composed of different sets of loci (Figure 2). We had multiple goals when separating loci into different matrices for phylogenetic analysis. First, we wanted to concatenate loci deemed to be of highest quality for our primary nuclear phylogenetic analysis. The 1,884 nuclear loci meeting the benchmarks for high quality (i.e., present in 25 or more species, with coding region) were concatenated together using Mesquite 3.02 as the concatenated nuclear matrix (total length = 2.41 Mb). Because concatenated datasets can overinflate support values (Kubatko and Degnan 2007), we also separated our concatenated nuclear matrix into 8 even subsets (nuclear subsets 1-8; 302,200 bp each) to better assess the strength of the phylogenetic signal. Contig order was randomized prior to concatenation and subset division to ensure that each subset represents a random sample of loci. Coding sites and non-coding sites of the concatenated nuclear matrix were also analyzed as two separate matrices (coding matrix, 1.93 Mb, and non-coding matrix, 565.8 kb).

We analyzed the remaining nuclear loci (those that do not meet all our "quality" requirements: present in less than 25 species, without coding regions), to see if these requirements truly mattered as indicators of quality and affect phylogenetic outcomes. There were 1,020 loci present in 15 to 24 species, and these were concatenated together as the missing species matrix (concatenated matrix length = 1.1 Mb). We did not assess loci present in less than 15 species because these were highly skewed towards species with much more sequencing data (*H. ophrys*, *H. festus, H. signatus, H. ustulatus*, *H.* sp. (SLCTY)). The 237 transcripts without coding regions present in 25 or more species were concatenated together as the unidentified transcripts matrix (118.1 kb).

The mitochondrial phylogeny includes an additional species, *H. paratus*, for which a DNA sample was sequenced rather than RNA. We only include the mitochondrial *H. paratus* sequences for phylogenetic analysis because these sequences are complete and have adequate coverage, while most nuclear sequences for *H. paratus* are not. Mitochondrial sequences for 16S RNA (1022 bp), 12S RNA (691 bp), ND1 (921 bp), ND2 (959 bp), ND3 (342 bp), ND4 (1289 bp), ND4L (268 bp), ND5 (1638 bp), ND6 (429 bp), ATP8 (158 bp), Cytochrome B (1111 bp), COX1 (1542 bp), COX2 (666 bp), COX3 (786 bp) were aligned, concatenated (concatenated mitochondrial matrix, 12.33 kb), and assigned codon positions based on annotations from the *H. oregonensis* mitochondrial genome (Masta and Boore 2004). To assess the strength of the mitochondrial phylogenetic signal with less data, we divided mitochondrial rRNA (1.72 kb) from protein-coding sites and then separated protein-coding alignment into 4 even subsets (mtDNA subsets 1-4; 2.53 kb each).

2.7.2. 16SND1 and Ef1-a phylogenies with broader species sample

To place our dataset in the context of the broader *Habronattus* phylogeny and previously sequenced data, sequence alignments from genes Ef1-a (819 bp) and 16SND1 (1047 bp) were integrated with pre-existing Sanger Sequencing data from Maddison and Hedin (2003). These datasets included a much broader sample of *Habronattus* species and populations. These combined Sanger and transcriptome alignments were reanalyzed using RAxML with 20 search replicates and a GTR+G+I substitution model. There were 160 Sanger sequences available for 16SND1 and 105 Sanger sequenced specimens for Ef1-a. The two Ef1-a introns (Intron 1 = 148 bp, Intron 2 = 91 bp) were coded as missing in transcriptome sequences. To see how additional species fell into the concatenated nuclear phylogeny while avoiding poor resolution of known clades due to low phylogenetic informativeness of the genes, any node recovered with minimum 90% bootstrap support in either the concatenated mitochondrial phylogeny (for the 16SND1 tree) or the concatenated nuclear phylogeny (for the Ef1-a analysis) was enforced as a phylogenetic constraint.

2.8 Introgression

2.8.1 Bayesian concordance analysis

Bayesian concordance analysis (BCA) uses a set of Bayesian gene trees to detect both the dominant phylogenetic signal arising from the genes, represented as a primary concordance tree (the estimated species tree), and secondary signals that are also substantially supported by some genes but discordant with the dominant species tree (Ané et al. 2007; Baum 2007). These discordant signals are considered to be part of the true evolutionary history of the group, and may be the result of introgression, incomplete lineage sorting (Larget et al. 2010), or noise due to

low phylogenetic informativeness of single genes (Smith et al. 2015). Genomic support for a particular clade is represented as a Concordance Factor (CF) and a 95% CF credibility interval. The CF represents an estimation of the proportion of sampled gene trees for which a clade is true (Baum 2007). The width of the credibility interval is indicative of the confidence of the CF. A wide credibility interval, and in particular an interval overlapping 0, indicates greater CF uncertainty due to lower posterior probabilities for that clade from Bayesian gene tree searches (Ané 2010)

We conducted all Bayesian gene tree searches with MrBayes 3.2.2 (Ronquist and Huelsenbeck 2003) with 4 chains (3 cold, 1 hot), 2 runs, and 25% burn-in. To determine the number of MCMC generations required for convergence, 20 genes were run until convergence as a test. Convergence was measured as a standard deviation (SD) of split frequencies <0.01 (Ronquist et al. 2005). The convergence time ranged from 570,000 to 13,087,000 and averaged 4,025,500 generations, so we set the number MCMC generations conservatively at 20,000,000 generations per gene. Codon positions were used as partitions.

BCA analyses were conducted using BUCKy 1.4.3 (Larget et al. 2010) with 2 runs and 4 chains per analysis. BUCKy does not assume that any specific biological process is causing gene tree incongruence (Larget et al. 2010) and so can be used to identify potential sources of introgression or incomplete lineage sorting. Due to the computational limitations of BUCKy, we only analyzed the *americanus* group and the VCC group, and only included genes longer than 1 kb that were present in all species being analyzed. There were 679 genes included for the *americanus* group analysis (with *H. signatus* outgroup) and 517 genes included for the VCC group analysis (with *H. ophrys* outgroup).

The single adjustable parameter in BUCKy is α, which represents the expected level of gene tree discordance; a larger value of α corresponds to greater gene tree discordance. To test α, we tried values of α = 0.1, 1, 2, 5 and 10. We found no substantial difference in results using different α for the *americanus* group, so kept α set to 1. The VCC clade had difficulty converging at higher α, so α was set to 0.1 for that analysis, although there was very little difference between CFs depending on α for VCC. The *americanus* group analysis ran for 10,000,000 generations and the VCC clade analysis ran for 30,000,000 generations due to the longer time required for convergence.

2.8.2 Patterson's D statistic and D_{FOIL}

To distinguish between incomplete lineage sorting (ILS) and introgression patterns in SNP data, we conducted Patterson's D statistic tests (Durand et al. 2011) and the related test D_{FOIL} (Pease and Hahn 2015). These tests compare patterns of shared SNPs across sets of 4 and 5 taxa, respectively. Under ILS in a 4-taxa binary tree, it is expected that the number of shared alleles between non-sister taxa would be equal for each possible non-sister pairing (i.e., patterns ABBA and BABA) (Durand et al. 2011). Introgression between two species is detected when they have significantly more shared alleles than expected (i.e., more ABBA than BABA or vice versa). The same principle can be applied to a 5-taxa tree, with some added complexity (Pease and Hahn 2015). Because there are more possible ways to share alleles and have introgression between 5 taxa, D_{FOIL} uses 4 different D statistics (one per species tested) to detect a broader range of possible events. Whether Patterson's D statistic or D_{FOIL} was used depended on the structure of the species tree of the taxa being tested. If the species tree did not agree with the

phylogenetic structure (((species 1, species 2),(species 3, species 4)), outgroup) required by D_{FOIL}, Patterson's D statistic was used instead.

Specimens tested (listed in Table 1) have some variability in sex chromosomes. We included female specimens (i.e., *H*. sp. (SLCTY), *H*. sp. (ESTU)), males with Y chromosomes (i.e., *H. borealis*, *H.mexicanus*, *H. zapotecanus*, and *H. decorus),* and specimens with have unknown karyotypes (i.e., *H*. sp. (BLNDI), *H*. sp. (ESTU), *H. clypeatus*, *H*. sp. (SLCTY)), because excluding all of these species from tests would impose too many limitations on testable hypotheses. An excess of shared alleles between species possessing Y chromosomes (or between species lacking a Y) resulting in a false positive for introgression remains a possibility, although the effect should not be substantial. *Habronattus* Y chromosomes are very young neo-Ys resulting from an X-autosome fusion that most likely retain some level of recombination with the neo-X (Maddison and Leduc-Robert 2013; Maddison 1982). As a result, they are assumed to be mostly, if not all, orthologous to autosomes of species lacking Ys. Any Y-specific gene should be excluded from the analysis, because only sites present in all 4 or 5 species were considered when counting SNPs.

We tested sets of species based on indications of introgression in previous studies (Maddison and Hedin 2003), in the mitochondrial phylogeny, or in the Bayesian concordance analysis. There were 4 principal hypotheses tested: (1) introgression among species of the *americanus* group, (2) introgression between the *clypeatus* group and *coecatus* group, (3) introgression between *H*. sp. (ROBRT) with the VCC group and (4) introgression between the *decorus*/*altanus*/*banksi* group and the *americanus* group.

For hypothesis 1, we tested all 5 *americanus* group species. For hypotheses 2, 3, and 4, we did not test all possible species because multiple large clades were involved. Instead, we selected

a few species as representatives of each clade. With a careful choice of species as a clade representative, we can still draw general conclusions about introgression across *Habronattus* clades. For hypothesis 2, we included 2 *coecatus* and 2 *clypeatus* species per D_{FOIL} test. *H. pyrrithrix* (*coecatus* group) and *H. clypeatus* (*clypeatus* group) are our clade representatives that are always included. *H. pyrrithrix* was chosen because its phylogenetic position is consistent in both the concatenated nuclear and mitochondrial phylogenies, and *H. clypeatus* was chosen because of mitochondrial introgression detected in the mitochondrial phylogeny and in other members of its species group by Maddison and Hedin (2003). By using more or less divergent sister taxa in different tests for comparison, we can better approximate where in the *coecatus* and *clypeatus* group introgression has occurred. For example, *H. pyrrithrix*/*H. mexicanus* tests for introgression across the coecatus group because *H. mexicanus* is a distant *coecatus* group species, *H. pyrrithrix*/*H. borealis* tests for introgression at a shallower scale, and *H. pyrrithrix*/*H.* sp. (BLNDI) tests for very recent introgression with either sister species. Likewise, *H. clypeatus*/*H.* sp. (SLCTY) tests for recent introgression with *H. clypeatus* or *H.* sp. (SLCTY) while *H. clypeatus*/*H. aztecanus* tests for introgression across the *clypeatus* group. For hypothesis 3, we tested only the species with the most sequencing data per clade for the four clades involved (*H.* sp. (ROBRT), and *viridipes*, *clypeatus, coecatus* groups). For hypothesis 4, we used the species with the most sequencing data for each clade involved, and to identify the depth of introgression, we used species at incremental phylogenetic distances from *H. decorus*.

D-statistics and D_{FOIL} tests are sensitive to missing data. If the phylogeny is sparsely sampled, these tests may detect introgression from a lineage that is not sampled (i.e., a ghost lineage) and incorrectly attribute the introgression signal to the lineage most closely related to the ghost taxon (Eaton et al. 2015). We only sampled 2 of 10 known *viridipes* species, 3 of 11 *clypeatus* group

species, 7 of 17 *coecatus* group species, and 5 of 11 *americanus* group species. As a result, even if we tested all possible combinations of sampled species, it is impossible to attribute an introgression signal to a specific species because there is a high probability that ghost taxa have had an effect (Eaton et al. 2015).

D_{FOIL} was run in mode dfoilalt to reduce noise from synapomorphic sites. We used a custom R script to count allele patterns. All sites that included ambiguous nucleotides, gaps, or missing data were excluded from the analysis. D_{FOIL} estimated divergence T-values were verified against the assumption that $T_{12}<T_{34}<T_{1234}$. We adjusted significance for 68 comparisons (Bonferroni correction) to a *p*-value lower than 0.00074 to indicate 95% significance.

3. Results

3.1 Transcriptome assemblies

Total raw sequencing reads, trimmed sequencing reads, and the number of reads assembled, and total transcripts, and coverage are summarized for every species in Appendix 2.

3.1.1 Reference transcriptome

The unfiltered *H. ophrys* reference transcriptome includes 117,859 transcripts (total 53,927,457 bases assembled), with an N50 (analogous to median contig length, Miller et al. (2010)) of 516, and an average coverage of 103x. Following filtering for redundancy, selection of a single variant per gene, removal of possible paralogs, and the separation of connected transcripts, there were 92,343 transcripts left. After reads were remapped, 51,143 transcripts had sufficient (5x) coverage (all filtering steps are summarized in Figure 1).

There are 13,901 *H. ophrys* loci with ORFs longer than 200 bp and coverage greater than 5x. Coding regions are most frequently identified in loci longer than 1000 bp (95%) and this number decreases the shorter the loci (500 to 1000bp: 77%, 200-500 bp: 15%). Almost all transcripts with coding regions (93%) have significant SWISS-PROT hits, compared to only 66% of the 78,442 contigs that do not have any detected ORFs. There are 18,053 *H. ophrys* loci with significant hits to the African social velvet spider protein set (26,889 proteins, 67.1% overlap) and 16,796 *H. ophrys* loci have significant hits with the Tarantula protein set (76,238 proteins, 22% overlap). These reference protein sets serve as best approximations of a complete spider transcriptome.

3.1.2. Reference-based transcriptomes assemblies

Reference-based transcriptome assemblies mapped on average 77% of trimmed reads to either the nuclear reference transcriptome (average nuclear coverage = 67x) or mitochondrial reference genome (average mitochondrial coverage = 13,640x). There were an average of 10,164 transcripts assembled per species, although numbers ranged widely (depending on the number of reads) from 3,746 for *H.* sp. (ROBRT) to 28,846 for *H. festus*. Contigs without identified coding regions were much more likely to be discarded based on high levels of ambiguous sites (59% in the transcripts without coding regions matrix compared to 11% in concatenated nuclear matrix and 22% in missing species matrix), and sequences discarded based on ambiguity levels were overwhelmingly short transcripts under 500 bp.

3.2 Substitution model selection

GTR+G+I or GTR+G was chosen as the optimal substitution model using AIC for all mitochondrial partitions. For the 20 nuclear genes tested, AIC selected the General Time Reversible (GTR) substitution model for 33 partitions; transversion model (TVM) for 22 partitions; unequal-frequency Kimura 3-parameter (K81uf) for 10 partitions; transition model (TIM) for 5 partitions; Hasegawa-Kishino-Yano (HKY) for 4 partitions; and Tamura-Nei (TrN) for 3 partitions. These include additional variations of these models, like gamma-distribution of rates (+G) or proportion of invariable sites (+I). We were unable to set a different model for each partition in concatenated matrices due to computational limitations. Instead, we set GTR+G+I as the substitution model in all phylogenetic Maximum Likelihood (ML) and Bayesian analyses (nst = 6 rates = invgamma) because it was the most commonly chosen and most widely applicable model.

3.3 Nuclear phylogenetic analyses

We resolved most phylogenetic relationships with high bootstrap support (95-100%) in the concatenated nuclear phylogeny (Figure 3). This phylogeny replicates clades previously resolved in the molecular phylogeny of Maddison and Hedin (2003) and the morphological phylogeny of Griswold (1987) with high bootstrap support. The *amicus*, *agilis*, and *tranquillus* groups are well supported and form the monophyletic and most basal AAT group, a pattern consistent with previous molecular and morphological phylogenetic studies (Maddison and Hedin 2003; Griswold 1987)**.** All *Habronattus* other than the AAT group (i.e., those at the bottom, or the right side of the phylogeny in Fig.3) also form a clade. Within this clade forming the right of the phylogeny, *H. geronimoi* is sister to the rest of the clade. The VCC group, and the *americanus* group are also resolved with high support.

The concatenated nuclear phylogeny also provides clarifications for several previously ambiguous phylogenetic relationships. Two well-supported clades are nested in the *coecatus* group: *H. festus*, *H. captiosus* and *H. borealis* together will be referred to as the Northern clade (all specimens in this clade were collected in Canada), while *H.* sp. (BLNDI) and *H. pyrrithrix* are together and form the Southern clade (since they are both found in the Southern USA neighboring Mexico). *H. virgulatus* is also sister to the Northern *coecatus* clade with high support (nuclear bootstrap = 97%). *H. altanus* with *H. decorus* are sisters, and the *banksi* group (*H.* sp. (CHMLA) and *H. zapotecanus)* is their sister. The previously intractable *H. hallani* is strongly supported as sister to the *pugilli*s group (*H.* sp. (SUNGL) and *H. pugillis)*. The relationships between the subgroups of the VCC clade are resolved with high bootstrap support; *H.* sp. (ROBRT) is sister to the *viridipes*, *clypeatus*, and *coecatus* group, the *viridipes* group (nuclear bootstrap = 83%) is sister to the *clypeatus* and the *coecatus* group, and the *clypeatus* and

coecatus group are sisters. *H. jucundus* groups with *H. calcaratus* rather than with the *oregonensis* group (a poorly supported relationship found in Maddison and Hedin (2003)), confirming that the *viridipes* group is in fact monophyletic. The internal relationships of the *americanus* group and the VCC group are also well resolved.

Most of the relationships found in the concatenated nuclear phylogeny are also replicated in phylogenies from different matrices (i.e., nuclear subsets 1-8, coding sites, non-coding sites, missing species, and transcripts without coding regions matrices). This suggests that high quality orthologs were assembled, even in matrices predicted to be of "lower" quality (i.e., the transcripts without coding regions matrix and missing species matrix), and that the phylogenetic signal for most branches is robust even with less data (i.e., nuclear subsets 1-8), and partitions that fit to different evolutionary models (i.e., coding sites, non-coding sites). The AAT, *americanus, altanus/decorus/banksi, pugillis*, and VCC groups are always monophyletic across analyses, as are the positions of *H. geronimoi* and *H. hallani*, and the internal relationships within the *americanus* group and the VCC clade are, for the most part (see below), consistent across analyses.

The most poorly resolved nodes of the concatenated nuclear phylogeny are also the most variable across the phylogenies from nuclear subsets 1-8, coding sites, non-coding sites, missing species, and transcript without coding regions (variable nodes are summarized in Figure 4). *Pellenes* cf. *levii* falls outside *Habronattus* (as expected) in the Maximum likelihood concatenated nuclear phylogeny, but falls within the *Habronattus* in most other nuclear phylogenies, including the bootstrapped consensus concatenated nuclear phylogeny (bootstrap support = 67%) (Fig. 4A-D). The positions for *H. icenoglei, H. oregonensis* and *H. cambridgei* also conflict across nuclear analyses (Fig. 4E-H). *H. icenoglei* and *H. oregonensis* are sisters in

the Maximum likelihood tree as well as in phylogenies from 3/8 nuclear subsets, coding sites, and non-coding sites (Fig. 4E). However, *H. cambridgei* and *H. oregonensis* are sisters in the bootstrap consensus tree (bootstrap support = 66%) and in phylogenies from 2/8 nuclear subsets (Fig. 4F). *H. icenoglei* is the most volatile species, even grouping at times with the VCC clade in 3/8 nuclear subsets (Fig. 4H).

There are other relationships that, despite relatively high bootstrap support, show some variability across phylogenies from different nuclear matrices (Fig. 5). *H.* sp. (ROBRT), while strongly supported as sister to the rest of the VCC clade (Fig. 5A), also groups frequently with the *clypeatus* group (Fig. 5B). *H. virgulatus* (bootstrap support = 97%) departs from its dominant concatenated nuclear position (Fig. 5D) to group with either the Southern *coecatus* clade (Figure 5E) or as a basal branch of the *coecatus* group (Fig. 5F) in some nuclear subsets. *H. ophrys* is positioned as the sister to *H. sansoni* and *H. americanus* in half of all data subsets (Fig. 5H), despite being sister to *H. tarsalis* (bootstrap support = 100%) in the concatenated nuclear phylogeny (Fig. 5G).

High bootstrap support (95-100%) should be indicative of good resolution for a phylogenetic relationship in the concatenated nuclear phylogeny. Nodes with high bootstrap support are replicated in most or all additional phylogenetic analyses using different matrices. Therefore, based on the criterion of high bootstrap support and consistency across analyses, we consider the concatenated nuclear phylogeny (Fig. 3) to be a reflection of the true species phylogeny, except for some ambiguity that remains regarding the positions of *Pellenes* cf. *levii* and *H. icenoglei*/*H. oregonensis*/*H. cambridgei* (Fig. 4). Because of high bootstrap support, we accept the positions of *H.* sp. (ROBRT), *H. virgulatus*, and *H. ophrys*/*H. tarsalis* (Fig. 5) despite variability found across analyses for their phylogenetic positions.

3.4 Mitochondrial phylogenetic analyses

The mitochondrial phylogenetic signal is strong for most clades and consistent across rRNA and protein-coding mitochondrial subsets (Figure 6). The *americanus* group and the VCC group remain monophyletic in all subsets. The most variability across subsets is found on the left side of the phylogeny (i.e., the outgroups, *H. paratus, H. geronimoi* and the AAT clade in Figure 3), although *H. oregonensis, H. hallani*, and the *pugillis* group also vary.

There are several notable differences between the mitochondrial and the nuclear phylogenies. Those most convincingly suggestive of introgressive hybridization involve distant, morphologically disparate species grouping together in the mitochondrial phylogeny. *H. clypeatus* is found with *coecatus* group species *H.* sp. (BLNDI), *H. borealis*, and *H. captiosus* with strong support (bootstrap support = 100%, replicated in each of rRNA and protein-coding subsets 1-4). This pattern of mitochondrial introgression repeats a pattern of *clypeatus*/*coecatus* discordance previously identified in Maddison and Hedin (2003), where *H. velivolus* (*clypeatus* group) and *H.* sp. (CHIH) (*clypeatus* group) specimens were polymorphic for *coecatus* and *clypeatus* group mitochondria. Unlike the nuclear phylogeny, *H. altanus, H. decorus,* and the *banksi* group, branch together at the base of the *americanus* group (bootstrap support = 100%, replicated in rRNA, protein-coding subsets 1-4). There are also mito-nuclear differences in the internal relationships of the *americanus* and *coecatus* groups. Because these species are very recently diverged, both ILS and hybridization are likely explanations for mito-nuclear discordance. Ongoing hybridization could explain why *H. ophrys* is the sister taxon to sympatric species *H. americanus* (bootstrap support = 100%), rather than to morphologically similar *H. sansoni* (sister to *H. ophrys* in the nuclear tree). Another difference between more closely related taxa is the relationship of *H. icenoglei* as sister to the *pugillis* group, *H. hallani,* and the VCC

group (bootstrap support = 86%) rather than *H. oregonensis*. There is also some support for the monophyly of *H. hallani*, *pugillis* group and VCC group (bootstrap support = 52%); however, the exact position of *H. hallani* within the clade is unresolved.

3.5 16SND1 and Ef1-a phylogenies (broader species sample)

When transcriptome and Sanger sequenced 16SND1 (Figure 7) and Ef1-a (Figure 8) sequences are combined and analyzed using ML methods, gene trees resolve most established species groups. In the 16SND1 tree, the *americanus* group and the *coecatus* group are each monophyletic, although they are poorly resolved at the species level. With the exception of *H. borealis*, *H. virgulatus,* and *H. festus,* specimens of the same species do not form individual clades. Instead, they sometimes group together and sometimes group with other species that are closely related. The transcriptome (GLR218) and Sanger (d436) *H.* sp. (SUNGL) specimens are not sisters but they do form a clade with two *H. pugillis* Sanger specimens (M-105MX, M-29GL), confirming the place of *H.* sp. (SUNGL) within the *pugillis* species complex. Both *clypeatus* group specimens *H.* sp. (CHIH)-HA292 and *H. velivolus*-HA659 are found within the *coecatus* group, replicating the patterns of mitochondrial introgression from the *coecatus* to the *clypeatus* group found in Maddison and Hedin (2003). *H. clypeatus*-GLR227 (transcriptome) is sister to *clypeatus* group *H.* sp. (CHIH)-HA292 (Sanger) within the *coecatus* group.

The Ef1-a tree resolves only some clades supported by the transcriptome concatenated nuclear phylogeny and is poorly resolved at the species level across the entire phylogeny. Only the outgroups, AAT group, *dorotheae* group, *americanus* group, and VCC group (with the exception of a few specimens) are monophyletic. The poor resolution and variable branch lengths in the Ef1-a tree can be explained by two factors. First, a short and highly conserved

alignment is at least partially responsible for the low resolution of the phylogeny. Second, there are high levels of ambiguous sites in the transcriptome sequences. The prevalence of ambiguous sites for the Ef1-a alignment is over 3%, and so this locus would have been discarded from the concatenated nuclear matrix. Such high levels of ambiguous sites may be the result of multiple paralogs or transcript variants assembled together on the reference and summarized as a single consensus sequence. Ambiguous sites are most concentrated in exon 3, and this exon's best BLAST hit is a *Salticidae* that is not *Habronattus*. The 1st and 2nd exon also have higher levels of ambiguous sites in the transcriptome specimens compared to the Sanger specimens, but their best BLAST hits are *Habronattus* Ef1-a. Even with the 3rd exon removed (as in the final alignment), transcriptome specimens have a tendency to group together, particularly in the "middle" portion of the phylogeny (i.e. groups *pugillis, hallani, altanus, decorus, banksi, fallax*). Unlike other transcriptome specimens, *H. festus*-GLR088, 094 does not have any ambiguous sites, resulting in a long branch for that specimen. The absence of ambiguous sites in only this sequence is further indication that multiple variants are most likely summarized as a single Ef1-a sequence. There are two Ef1-a variants known so far, one with 3 exons (Sanger specimens included here), and another with 2 exons (Hedin and Maddison 2000). The concentration of ambiguous sites in the 3rd exons suggests that other variants may exist.

3.6 Introgression

3.6.1 Bayesian concordance analysis

We only report CFs with credibility intervals that do not overlap 0 (and as a result are considered to be significant, Ané 2011). We only included CFs with credibility intervals overlapping 0.1 or higher for inclusions in Figures 9 and 10. CFs with credibility intervals

overlapping values as low as 0.05 are also listed in the Appendix 3 (*americanus* group) and 4 (*coecatus* group). Lower CFs (under 0.05) that did not overlap 0 were also prevalent, although we do not report them all because low CF values are likely to be overestimated (Ané 2011) and could be especially vulnerable to false positives due to uninformative genes causing noisy data.

americanus group – The BUCKy analysis of 679 *americanus* group loci (105 tree topologies sampled, 25 distinct splits) converged with an average SD of mean sample-wide CF of 3.24×10^{-5}. The primary concordance phylogeny (Figure 9) reflects the concatenated nuclear phylogeny and is well supported by most gene trees. There is a number of secondary CFs supported by a large portion of the data. *H. ophrys* is linked to *H. americanus* and *H. sansoni* with strong support (CF = 0.272 - 0.339). This high CF is more than twice the CF shared for the equivalent pairing *H. tarsalis* and *H. sansoni*/*H. americanus* (CF = 0.112 - 0.162), suggesting that ILS and shared ancestry can only account for approximately half of that shared genetic history. *H.* sp. (ESTU) is substantially linked with *H. tarsalis* (CF = 0.112 - 0.162). However, it is not significantly linked (CF = 0) with *H. ophrys,* another indication of possible introgression. In addition to those clades that are listed in Appendix 3, there are 12 clades with CFs under 0.05 that are significant (i.e., they do not overlap 0) but which are not listed.

VCC group - The BUCKy analysis for 517 VCC group loci (4,795,750 tree topologies sampled, 8,177 distinct splits) converged with an average SD of mean sample-wide CF of 0.003. The longer time to convergence and lower CFs overall compared to the *americanus* group could be indicative of a higher level of discordance in the VCC group. Low informativeness of genes could also explain low CFs and difficulties with convergence, as well as the large number (4,795,750) of tree topologies sampled. The primary concordance phylogeny (Figure 10) reflects the concatenated phylogeny except for the positions of *H. virgulatus*, which is a part of the

Southern *coecatus* clade (i.e., *H. pyrrithrix* and *H.* sp. (BLNDI)) rather than the Northern clade in the primary concordance tree. *H.* sp. (ROBRT) has two strong secondary CFs: it is grouped with the *clypeatus/coecatus* clade (CF = 0.174 – 0.222) and with the *clypeatus* group (CF = 0.17 – 0.205). Support linking *H.* sp. (ROBRT) to the *clypeatus* group is unanticipated based on their phylogenetic distance, and is unlikely to be due entirely to ILS given a very low CF (0.039 – 0.06) with the equidistant *viridipes* group. There is some support (CF=0.087 – 0.122) that ties *viridipes* and *coecatus* together as a group. The species of the *coecatus* group have many secondary CFs amongst them, possibly because of widespread ILS. *H. virgulatus* has a low primary CF (CF=0.101 – 0.137) and 7 secondary CFs > 0.05 within the *coecatus* group, the most genetic discordance of any species. *H.* sp. (BLNDI) and *H. pyrrithrix* are the other *coecatus* species most often implicated in secondary signals. In addition to those clades that are listed in Appendix 4, there are 181 clades with CFs under 0.05 that are significant (i.e., they do not overlap 0). Of these 181 clades, 35 clades with CFs below 0.04 (average near 0.01) group together species from the *clypeatus* and the *coecatus* group.

3.6.2 Patterson's D statistic and D_{FOIL}

Allele counts are listed for D_{FOIL} and D-statistics in Appendices 5 and 6. D values and *p* values are reported in Appendices 7 and 8.

(1) Among species of the americanus group (Figure 11)

Directional introgression is detected twice within the *americanus* group: from the common ancestor of *H. sansoni* and *H. americanus* to *H. ophrys* (Figure 11A, D_{FO}=0.267 $p < 10^{-12}$, D_{IL}=0.235 p=0, D_{FI}=0.019 p=0.737, D_{OL}=0.075 p=0.178) and between *H.* sp. (ESTU) and *H. tarsalis* (Figure 11B, D= 0.146, $p < 10^{-12}$)

(2) Between the coecatus and the clypeatus group (Figure 12)

Introgression is detected only once between the *coecatus* group and the *clypeatus* group. It is detected between *H. aztecanus and H. clypeatus* and *H. pyrrithrix* when the fourth species included is the distant *coecatus* group *H. mexicanus* (Figure 12C, D_{FO}=0.146 $p < 10^{-12}$, D_{IL}=0.155 $p < 10^{-12}$, D_{FI}=-0.015 p=0.634, D_{OL}=-0.026 p=0.391). The same signal is weakened due to a D_{FO} = 0 when the less distant *coecatus* group species is *H. borealis (*Figure 12F, D_{FO}=*-0.094 p=0.001*, D_{IL}=*-0.13164 p=4 x* 10^{-6}, D_{FI}=*-0.011 p=0.7334*, D_{OL}=*-0.055 p=0.07757)*. This signature (D_{FO}=0 D_{IL}=- D_{FI}=0 D_{OL}=0) is not indicative of any introgression, although it does hint at introgression between *H. clypeatus* and *H. pyrrithrix*. Similarly, reciprocal introgression is detected between *H. clypeatus* and *H. pyrrithrix* when species 1 is *H.* sp. (SLCTY) and species 4 is either *H. mexicanus* (Fig.12D, D_{FO}=-0.096 p=0.003, D_{IL}=-0.167 $p < 10^{-12}$, D_{FI}=-0.011 p=0.814, D_{OL}=-0.166 $p < 10^{-12}$) or *H. borealis* (Fig12E, D_{FO}=-0.0079 p=0.802 D_{IL}=-0.14145 p=6 x 10^{-6} D_{FI}=0.113 p=0.01104 D_{OL}=-0.157 p=0.00043). Support for introgression between *H. pyrrithrix* and the *clypeatus* group disappears when *H.* sp. (BLNDI) is the fourth species (Fig12A-B), ruling out *H. pyrrithrix*-specific introgression. The signal is most likely from a ghost taxon from the Southern clade that is closely related to *H. pyrrithrix*.

(3) Between H. sp. (ROBRT) and the VCC clade (Figure 13)

H. sp. (ROBRT) may have hybridized with both the *coecatus* and the *clypeatus* group. Introgression is detected between *H.* sp. *(*ROBRT*)* and *H.* sp. (SLCTY) when the third species i*s H. festus (*Fig.13B D=0.360, $p < 10^{-12}$*)* or *H. jucundus (*Fig.13C, D=-0.405, $p < 10^{-12}$*)*. Introgression is also detected between *H. festus* and *H.* sp. (ROBRT) when the *clypeatus* group is excluded *(*Fig.13D, D=-0.166, $p < 10^{-12}$*)*.

(4) Between H. decorus and the americanus group (Figure 14)

Introgression is detected with both D-statistics (Fig.14A, D= 0.17 $p < 10^{-12}$) and D_{FOIL} (Fig.14A, D_{FO}=-0.164 $p < 10^{-12}$, D_{IL}=-0.116 $p < 10^{-12}$, D_{FI}=-0.004 p=0.883, D_{OL}=0.117 $p < 10^{-12}$) between *H. decorus* and *H. ophrys* when compared to *H. zapotecanus*. When the sister to *H. decorus* is a more distant species (*H. cambridgei, H. oregonensis, H. jucundus, H. festus*), the introgression signal between *H. decorus* and *H. ophrys* disappears (Fig.14B-E), and D-statistics introgression signals emerge instead between *H. ophrys* and the more distant species *H. oregonensis* (Fig.14C, D= -0.11 p=1.6 x 10^{-11}), *H. jucundus* (Fig.14D, D=-0.15 $p < 10^{-12}$), and *H. festus* (Fig.14E, D=-0.1 p=2 x 10^{-12}). Introgression is also detected between *H. oregonensis* and the *americanus* group (*H. ophrys*/*H.* sp. (ESTU)) using D_{FOIL} (Fig.14C, D_{FO}=0.104 $p < 2$ x 10^{-10}, D_{IL}=0.111 $p < 10^{-11}$, D_{FI}=0.068- p=0.008, D_{OL}=0.085 $p < 9$ x 10^{-4}). D_{FOIL} tests do not detect introgression between *H. ophrys* and *H. jucundus* or *H. festus*. These D_{FOIL} signatures are inconsistent with a lack of introgression (expected signature: D_{FO}=0 D_{IL}=0 D_{FI}=0 D_{OL}=0), and they could be caused by multiple introgressions (D_{FO}=+ D_{IL}=+ D_{FI}=0 D_{OL}=+). Unexpected introgression detected between the *americanus* group and distant species *H. festus*, and *H. jucundus* (represented by D_{FO}=+ D_{IL}=+) may conceal introgression between *H. decorus* and *H. ophrys* detected using D-statistics (represented as the positive D_{OL}=+).

4. Discussion

Using transcriptomes, we were able to reconstruct a phylogeny for 34 species of *Habronattus* with high confidence at most nodes using 1,884 orthologous loci. While previous phylogenies for *Habronattus* were poorly resolved and limited to a few mitochondrial and nuclear genes (Maddison and Hedin 2003), this improved phylogeny will permit further research on character evolution, sexual selection, and hybridization, within a robust phylogenetic framework. We also integrated Ef1-a and 16SND1 transcripts with Sanger sequencing data from Maddison and Hedin (2003), clarifying some of their results and placing our data in the context of a more complete phylogeny. Based on differences in the mitochondrial and nuclear phylogenies, we infer several possible instances of mitochondrial introgression. We identified extensive nuclear genetic discordance indicative of ILS or introgression within the VCC group and the *americanus* group using BCA, and we identified several instances of nuclear introgression across the transcriptome phylogeny using D-statistics and D_{FOIL}.

4.1. *Habronattus* phylogeny

The concatenated nuclear phylogeny is generally in agreement with previous work, finding strong support for the monophyly of the AAT, VCC, and *americanus* group (Maddison and Hedin 2003; Griswold 1987). The positions of *H. hallani*, *pugillis* group, *altanus*/*decorus*/*banksi* groups, *H.* sp. (ROBRT), as well as the internal relationships of the VCC group and the *americanus* group, are also clarified with this phylogeny. The VCC clade has very low dominant CFs in the BCA primary concordance tree, potentially indicating that the group is still in the early stages of divergence with widespread incomplete lineage sorting and possibly also ongoing hybridization. The *americanus* group on the other hand, has stronger genetic concordance (high

dominant CF values) even though it is also very recently diverged (BEAST analysis, Maddison and Leduc-Robert 2013).

There is only mild support for the monophyly of *Habronattus* in the concatenated nuclear phylogeny, and there is considerable variation in the position of *Pellenes* across analyses. The possibility of non-monophyly of *Habronattus* has been previously raised based on similarly variable positions for *Pellenes* specimens within a set of phylogenetic analyses (Maddison and Hedin 2003). However, non-monophyly of *Habronattus* is unexpected, because the group is well defined morphologically. For instance, the elbowed tegular apophysis (a component of the male palp), a trait unique to *Habronattus* and shared by most species (with the exception of secondary loss in the *coecatus* group), is not found in any *Pellenes* species. This trait appears to be a synapomorphy consistent with monophyly of the genus; however, the molecular boundary delineating *Habronattus* and *Pellenes* was difficult to resolve with the specimens we sampled. *Pellenes* cf. *levii*-GLR106 (transcriptome), a North American specimen, when it is integrated with 16SND1 data from additional species (Fig. 7), pairs with the ingroup *Pellenes* specimens (Sanger specimens: HA093, HA510, HA430, HanMN105, MAA2005), also from North America. These same *Pellenes* (Sanger) specimens are outgroups in the Ef1-a phylogeny (Fig. 9). Because *Pellenes* are also distributed in Asia (Logunov 1999) and Europe (Fiser and Azarkina 2005; Buchholz 2007), a broader global sample of *Pellenes* specimens and other closely related groups (e.g., *Harmochirus*) should be included to better tease apart relationships at the base of the *Habronattus* tree and determine if the genus is monophyletic.

With a better-resolved phylogeny, we can more accurately map Y chromosome origins onto the *Habronattus* phylogeny. The ancestral sex chromosome state for *Habronattus* species is XXO♂/XXXX♀ (Maddison 1982). An X-autosome fusion producing an XXY♂/XXXX♀

karyotype, and a second autosome-autosome fusion producing an XXXY♂/XXXXXX♀ karyotype, are also found in multiple species (Maddison and Leduc-Robert 2013). We reduce the maximum number of origins from 8 to 15 (Maddison and Leduc-Robert 2013) to 8 to 14. Because there is strong evidence that the *altanus* (XXY♂) and *decorus* (XXXY♂) groups are sisters and that *altanus*/*decorus* group is sister to the *banksi* group (XXY♂), we infer a single origin of Y chromosomes for this group and an additional fusion in the *decorus* group to yield an XXXY♂ from XXY♂ (reducing the total from three possible to two possible fusions). *H. hallani* (XXY♂) is sister to the *pugillis* group (XXO♂ and XXY♂); thus, we can either infer a single origin of Y in *H. hallani,* followed by a loss in *H. pugillis,* or two independent origins with no losses. Other inferred origins are unchanged by the new phylogeny; we still count 2 to 4 origins in the *coecatus* group, 1 to 3 origins in the *viridipes* group, one origin in *H.* cf. *paratus*, one in *H.* sp. (MACHAL), and a single origin in the *dorotheae* group.

4.2. Introgression in *Habronattus*

We examined nuclear and mitochondrial phylogenetic signals discordant to the species tree (the concatenated nuclear phylogeny, Fig. 3) for signs of possible introgressive hybridization. Because the mitochondrial genome is haploid and maternally inherited, mitochondrial introgression signals are undiluted by recombination from backcrossing and carry signatures of directional introgression from important processes such as sex-biased hybridization (Avise 2001). To assess nuclear genetic discordance for introgression, we first documented nuclear signals discordant with the species tree using a BCA, and then explicitly tested for introgression using D-statistics. BCA is a good first step to assess the extent of genetic discordance and the degree of divergence between species (Ané et al. 2007; Baum 2007) insofar

as it quantifies all of nuclear gene discordance from both ILS and hybridization. This information can be useful to interpret introgression in the context of overall discordance, and to estimate how much of the genome may have introgressed (estimated as a CF credibility interval; Baum 2007). Alternatively, D-statistics explicitly test for introgression and can determine directionality. D-statistics are also more sensitive to weak signals of introgression because they are based on shared alleles rather than gene trees. Gene trees, especially from genes with low informativeness due to low variability (a problem arising when working with closely related species like *Habronattus*), may create considerable noise in the BCA resulting in wider CF credibility intervals and possibly overestimated low CF values (Chung and Ané 2011). Only a small fraction of orthologs meet the criteria for gene tree construction and inclusion in the BCA (i.e., they are long enough to be informative and present in all species), while D-statistics can include all nuclear sequencing data because it does not depend on a gene tree construction step. Simulations indicate that introgression as low as 1% of the genome can yield positive D-statistic results (Good et al. 2015). Consequently, relative to BCA, D-statistics are expected to better detect ancient introgression that may result in only short introgressed genomic regions (Sankararaman et al. 2012) and in infrequent hybridization that may have not resulted in widespread nuclear introgression. Such low levels of introgression may be overinflated as CFs in the BCA, obscured during the gene tree construction process, or not sampled at all using a gene tree approach.

D-statistics, while highly sensitive to introgression, can also be skewed by other factors that are difficult to account for like multiple introgression events, ancient population structure, and selection (Durand et al. 2011). Multiple introgression events could reduce the power of the test if introgression occurred between sister taxa (Durand et al. 2011), or if multiple introgression

events are detected together and conflict (Pease and Hahn 2015). Patterns identical to introgression can be produced as a result of ancient population structure (Slatkin and Pollack 2008), and possibly also selection (Durand et al. 2011). A lineage with a much higher substitution rate might increase the likelihood of shared alleles with distant species (Pease and Hahn 2015). Lineages undergoing parallel selection in similar environments could also have a greater than expected number of shared alleles as a result of convergence. A larger sample of loci should reduce the effects of selection, since selection differences are unlikely to be substantial across the genome.

Because D-statistics are the only analyses to explicitly test for introgression, ILS cannot be ruled out as the cause of a discordant signal when only mito-nuclear discordance and BCA results is used to assess discordance. However, there are several indications that introgression may be the cause of a discordant signal rather than ILS without explicit testing. We infer that the mitochondrial grouping of morphospecies that are phylogenetically distant in the species tree is strongly suggestive of introgression because the likelihood of ILS is decreased with increasing phylogenetic distance due to deeper coalescence times (Pamilo and Nei 1988). Sympatric species grouping together in the mitochondrial but not the nuclear phylogeny may also be an indication of ongoing hybridization (Funk and Omland 2003). BCA CF credibility intervals can also be compared to determine if they are equivalently supported as expected under ILS, or if one contradictory clade has a greater CF, as expected under introgression. Non-overlapping CF credibility intervals for two contradictory secondary CFs can be thought of as comparable to an excess of ABBA or BABA pattern resulting in a D-statistic significantly different from 0.

By focusing more broadly on hybridization across the *Habronattus* phylogeny, and by using sensitive detection methods such as D-statistics, we were able to determine whether

introgression occurs between distantly related *Habronattus* clades. Most hints of hybridization detected in *Habronattus* prior to our analysis were restricted to closely related species and subspecies (*amicus* group - Hedin and Lowder 2009, *coecatus* group - Maddison and Hedin, unpubl., the *pugillis* group - Maddison and McMahon 2000, *H. americanus* - Maddison and Blackburn 2015), with the exception of more distant mitochondrial introgression detected between the *clypeatus* group and the *coecatus* group (Maddison and Hedin 2003). We detected both introgression between closely related *americanus* group species and between several more distantly related clades (e.g., *clypeatus* group and *coecatus* group, *H.* sp. (ROBRT) and the *clypeatus* group, *americanus* group and *H. decorus)*. These results are unexpected based on levels of divergence in courtship behaviour and morphology between clades. Nuclear introgression did not always co-occur with possible mitochondrial introgression. Significant admixture highlighted by both the BCA and D-statistics was almost exclusively detected in closely related species that are currently sympatric, while more ancient introgression signals between species that do not share a contact zone had more minimal nuclear introgression better captured by D-statistics.

4.2.1. Introgression within species groups

americanus group - We identified two instances of introgression in the *americanus* group using D-statistics: between *H.* sp. (ESTU) and *H. tarsalis* and between *H. ophrys* and *H. americanus*/*H. sansoni* (Fig. 11). The BCA (Fig. 9) identifies both of these instances of introgression as substantial CFs as well. While there is no obvious co-occurring mitochondrial introgression for *H.* sp. (ESTU) and *H. tarsalis*, there is mitochondrial introgression detected between the sympatric *H. ophrys* and *H. americanus*. They are grouped together in the

mitochondrial phylogeny (Fig. 6, 100% mitochondrial bootstrap support), but not in the nuclear phylogeny (Fig. 3, 100% nuclear bootstrap support). D_{FOIL} tests indicate that introgression occurred between *H. americanus*/*H. sansoni* and *H. ophrys* (rather than *H. ophrys* and *H. americanus* only), and this is in accordance with BCA results which link *H. americanus*/*H. sansoni* and *H. ophrys* together (CF = 0.272-0.339). These results could be explained by introgression prior to *H. sansoni*/*H. americanus* speciation, or by multiple introgressions (between *H. sansoni* and *H. ophrys* and between *H. americanus* and *H.ophrys*). Introgression could also explain why *H. ophrys* and *H. americanus* are resolved as sister taxa by many nuclear data subsets (Fig. 5).

Morphological and behavioural differences in the *americanus* group are relatively few compared to other clades like the *decorus* group (Maddison and Hedin 2003). While each *americanus* group species has a distinct courtship display, they all have similar visual, acoustic and vibratory elements (W.P. Maddison, pers. comm.). *Americanus* group male palps are also distinctive in that the tibial apophysis is bifurcated, and the embolus and tegular apophysis are consistent in length and shape (Griwold 1987, Fig. 184-186). Among *americanus* group species, male ornaments exhibit the most variation. For example, ornaments range in colour from green (*H. ophrys*) to blue and red (*H. americanus*). It is possible that introgression could have influenced the rapid evolution of sexually selected male *americanus* group ornaments. For example, tufts above the front pair of eyes, a male ornament unique to the *americanus* group, is found only in *H. sansoni* and close relatives (*H. kubai,* some *H. americanus* populations), and more distant *H. ophrys*, with which our data suggests a shared history of hybridization.

coecatus group - There are a total of 15 secondary CFs greater than 0.05 detected within the *coecatus* group (and many more under 0.05), indicating high levels of nuclear discordance within the clade. This number would likely increase with a complete sample of *coecatus* group species (we only sampled 7 of 16 described species). Because the origin of the *coecatus* group is predicted to be very recent (see BEAST analysis, Maddison and Leduc-Robert 2013), ILS and hybridization are both likely contributors to the high levels of discordance detected. ILS cannot be ruled out as the only cause of discordance because we did not explicitly test for introgression using D-statistics. However, there are non-overlapping secondary CFs for contradictory clades, indicating that introgression may have occurred. For example, there is a significant CF linking *H. festus* and *H. captiosus* (CF = 0.130–0.172) and no significant CF for the contradictory clade *H. festus* and *H. borealis*, which could be indicative of introgression between *H. festus* and *H. captiosus*. There are also several indications of possible mitochondrial introgression. Pairs *H.* sp. (BLNDI)/*H. captiosus*, and *H. mexicanus*/*H. pyrrithrix* are sister taxa in the mitochondrial phylogeny, but not in the concatenated nuclear phylogeny.

It is difficult to say if courtship behaviour and morphology is divergent enough between *coecatus* group species to completely prevent hybridization. Short branch lengths in the 16SND1 tree and low primary CFs in the BCA indicate low genetic divergence within the group. Morphological differences are concentrated in male ornaments, which vary widely in colours and styles, while *coecatus* group genitalic morphology is more consistent across the clade (Griswold 1987). *Coecatus* group species all share the loss of an elbowed tegular apophysis, and all species have similar lengths, shapes, and rotations of the embolus and tibial apophysis (Griswold 1987). Courtship is very complex and unique in each species, although they are all variants of a display composed of similar motifs of display elements performed in a similar order (Elias et al. 2012).

The possibility of widespread hybridization within the *coecatus* group despite these differences should be explicitly tested using D-statistics with a complete *coecatus* group phylogeny.

4.2.2. Distant introgression

coecatus and clypeatus group - There is a strong signature of mitochondrial introgression between the *coecatus* group and *H. clypeatus* (Fig. 6), as well as a corresponding nuclear signal of introgression detected by D_{FOIL} (Fig. 12). D_{FOIL} results indicate that introgression must have occurred more recently than the *H. borealis*/*H. pyrrithrix* split within the *coecatus* group, and therefore does not pre-date divergence of the *clypeatus* and *coecatus* clades. *H. clypeatus* also falls within the *coecatus* group (rather than the *clypeatus* group as expected) in the concatenated mitochondrial phylogeny with 100% bootstrap support. There are some very low and significant CFs that also link *H. clypeatus* with *coecatus* group species together in a clade: *H. clypeatus*/*H. mexicanus* (CF = 0.006-0.017), *H. clypeatus*/*H. virgulatus* (CF = 0.004 - 0.017), *H. clypeatus*/*H. pyrrithrix* (CF = 0.002 - 0.012). In addition, there are 32 other CFs averaging 0.01 link species from the *clypeatus* group and *coecatus* group. While ILS or noisy gene trees could explain low-level discordance, some of these CFs could be a result of introgression. Patterns of mitochondrial introgression detected from *H.* sp. (CHIH)-HA292 and *H. velivolus*-HA659 in Maddison and Hedin (2003) are also replicated in the 16SND1 tree (Fig. 7). *H.* sp. (CHIH)-HA292 (*clypeatus* group member) falls within the *coecatus* group (and *H. clypeatus*-GLR227 is its sister), while the other specimen of *H.* sp. (CHIH)-HA272 falls clearly within the *clypeatus* group. *H. velivolus*-HA659 (*clypeatus* group) falls within the *coecatus* group as well, while *H. velivolus*-HA661 falls within the *clypeatus* group.

These results, coupled with those of Maddison and Hedin (2003), could be an indication of frequent and recent introgression between the *clypeatus* and *coecatus* group. This is unexpected because these clades differ substantially and consistently in morphology and courtship behaviour (Elias et al. 2012; Elias, Mason, and Maddison 2003; Maddison and Hedin 2003). Both groups have highly complex ornamentations and courtship displays. The *coecatus* group courtship display is composed of more stages, and its motions are different from those of the *clypeatus* group (Elias et al. 2012). For example, early stages of the courtship display have the *clypeatus* group moving with their front legs oriented downward, while the *coecatus* group keep their legs up (W.P. Maddison, pers. comm.). Vibratory and acoustic signals have been shown to be very important components of the display (Elias et al. 2003), and these signals are also very different between the two groups (Elias et al. 2003; Elias et al. 2012). Genitalia morphology also differ considerably between the two groups. *Coecatus* group species all share the loss of an elbowed tegular apophysis that is present in all *clypeatus* group species (Griswold 1987). The tibial apophyses of the *clypeatus* group are thinner and hook-like while they are thicker and more triangular in the *coecatus* group. The female epigyna of each group also have correspondingly different forms (Griswold 1987; W.P. Maddison, pers. comm.), consistent with a lock and key coevolutionary process. The possibility of hybridization despite these substantial clade differences is worth noting. Because of an evolutionary lag between male sexually selected traits and female preferences (Schluter and Price 1993), species-specific differences in male courtship traits do not always result in behavioural isolation. However, *clypeatus* group females are very likely to have evolved preference to *clypeatus* rather than *coecatus* group displays (and vice versa). Introgression between these two groups strongly suggests that distant hybridization is

possible despite morphological and behavioural differences and could have resulted in introgression of distant genes.

clypeatus group and H. sp. (ROBRT) - D-statistics indicate that the *clypeatus* group has a strong signal of nuclear introgression with *H.* sp. (ROBRT) (Fig. 13). This signal is also reflected in a larger CF for the *clypeatus* group/*H.* sp. (ROBRT) clade (CF = 0.1705-0.205) compared to the *coecatus* group/*H.* sp. (ROBRT) clade (CF = 0.027-0.058). Introgression detected between *H.* sp. (ROBRT) and the *coecatus* group is also detected using D-statistics. However, because introgression with the *coecatus* group is only detected in the absence of the *clypeatus* group, and because of the large difference in CFs found by the BCA, the introgression signal detected between *H.* sp. (ROBRT) and the *coecatus* group is most likely a result of introgression with the unsampled *clypeatus* group, which is acting a ghost lineage. There is not a clear and comparable mitochondrial introgression signal linking *H.* sp. (ROBRT) and the *clypeatus* group. Instead, *H.* sp. (ROBRT) is sister to the *clypeatus* and *coecatus* group in the mitochondrial phylogeny, and all four *H.* sp. (ROBRT) specimens in the 16SND1 tree fall within the *coecatus* group (Fig. 7). The BCA found substantial genomic support for *H.* sp. (ROBRT)/*clypeatus* group/*coecatus* group as a clade (CF = 0.174-0.222) compare to the contradictory clade *H.* sp. (ROBRT)/*viridipes* group, which has much less genomic support (CF = 0.039 – 0.068). ILS, introgression with an ancestral coecatus group/clypeatus group lineage, and introgression with the *clypeatus* group only, could all have contributed to these estimated CFs. More ancestral introgression could explain why there is strong genomic support (large CFs) for *H.* sp. (ROBRT) and the entire *clypeatus* group, or *clypeatus*/*coecatus* group, rather than with one or a few particular *clypeatus*/*coecatus* species specifically. However, the possibility of introgression in

ancestral lineages that pre-date divergence of the *clypeatus* and *coecatus* group clade requires further testing. It is also possible that denser sampling within the *clypeatus* group could uncover a lineage with recent or ongoing introgression with *H*. sp. (ROBRT).

The phylogenetic position of *H*. sp. (ROBRT) has been historically difficult to resolve based on molecular data, morphology, and courtship behaviour (Maddison and Hedin 2003). It is similar to both the *viridipes* and *clypeatus* groups in many ways, and introgression could be in part responsible for this unique combination of *viridipes* and *clypeatus* group traits. *H*. sp. (ROBRT) has a raised-V setae ridge on the male carapace and both 1st and 3rd legs are modified like the *viridipes* group (Maddison and Hedin 2003). On the other hand, its ventral and dorsal abdominal stripes are very similar to the *clypeatus* group (W.P. Maddison, pers.comm.). Males from some populations of *H*. sp. (ROBRT) have red-purple bumps on their 3rd legs, a trait found otherwise only in *clypeatus* group species *H*. sp. (CHIH), *H. formosus,* and *H. velivolus* (a species sympatric with *H*. sp. (ROBRT)) *(*W.P. Maddison, pers. comm.). *H*. sp. (ROBRT) and *clypeatus* group males (*H. velivolus, H. aztecanus*) also share a unique zig-zag pattern emerging from within the frontal eyes (W.P. Maddison, pers.comm.).

americanus group and distant clades - Based on phylogenetic distance, mitochondrial and nuclear introgression detected between the *americanus* group and the *decorus/altanus/banksi* group, as well as nuclear introgression with *H. oregonensis*, *H. jucundus*, and *H. festus*, were most unexpected. There is a strong signal of distant mitochondrial introgression between the *americanus* group and the *decorus/altanus/banksi* group, which together form a clade in the mitochondrial transcriptome phylogeny (Fig. 6) and the 16SND1 phylogeny (Fig. 7) but are not a clade and distant from each other in the nuclear phylogeny (Fig. 3). It is unlikely that ILS is the

cause of this topology, because the common ancestor of *H. ophrys* and *H. decorus* is relatively ancient and because the phylogenetic signal is strong across different mitochondrial data subsets. Moreover, D-statistics show a weak signal of nuclear introgression between *H. decorus*, but not *H. zapotecanus* (*banksi* group) (Fig. 14). Ancient introgression at the base of the *americanus* group with the *decorus*/*altanus*/*banksi* group may have resulted in negligible nuclear introgression and fixation of a single mitochondrial type, which was followed by diversification of extant species groups. Because introgression would have been ancestral to the diversification of these groups, the ancestral hybridizing *decorus* group and the *americanus* group members may have looked and behaved quite differently from extant species. However, extant *americanus* group and *decorus* group species have very different morphology (Griswold 1987) and behaviour (W.P. Maddison, pers. comm.) that would predict reproductive isolation. For instance, the *americanus* group male palp has an unrotated embolus and a thin bifurcated tibial apophysis, while *decorus* group palps have thicker and more triangular tibial apophyses and a rotated embolus (Griswold 1987).

The *americanus* group and the *coecatus* group, *clypeatus* group, and *oregonensis* group are even more genetically, morphologically, and behaviourally different (Griswold 1987, Maddison and Hedin 2003), yet introgression between these groups also emerged from D-statistics and D_{FOIL} tests. While it is not out of the question that introgression has frequently occurred between the *americanus* group and the rest of the *Habronattus* clade, signals of introgression between *H. ophrys* and *H. oregonensis, H. jucundus,* and *H. festus* could all be the result of a single ghost lineage belonging to a clade distant to *H. ophrys*. It is also possible that introgression occurred in an ancestral lineage prior to clade diversification. More extensive tests

for distant introgression would be required to clarify the specific origin and timing of these signals.

4.2.3. Potential drivers of hybridization

Our results suggests that distant introgression in *Habronattus* may be much more pervasive than we initially predicted based on divergent male traits. We not only found introgression between the *clypeatus* and *coecatus* groups (like in Maddison and Hedin (2003)), but also between *H.* sp. (ROBRT) and the *clypeatus* group, and between the *americanus* group and distant species. Sexual selection is an important process driving diversification and the evolution of male traits in these *Habronattus* groups (Masta and Maddison 2002). These differences in courtship behaviours and male ornaments are expected to result in behavioural reproductive isolation because they should generally coevolve with female preferences for those traits (Panhuis et al. 2001).

Directionality of introgression, determined by D_{FOIL} and the mitochondrial phylogeny, hint at what processes may be driving hybridization despite divergent morphology and behaviour. *H. clypeatus* falls within the *coecatus* group in the mitochondrial phylogeny, and several species of the *clypeatus* group are polymorphic for *coecatus* group mitochondria in Maddison and Hedin (2003). Nuclear introgression however is reciprocal. Directionality of mitochondrial introgression but not nuclear introgression could be the result of sex-biased hybridization, driven by a difference in female discrimination (Avise 2001), causing *coecatus* females to sometimes hybridize with *clypeatus* group males but not the other way around. *H. pugillis* females have been shown to sometimes choose foreign males from divergent populations (Hebets and Maddison 2005), a preference that could be driven by a bias for vibratory signals

that are more complex (Elias 2006). This could be a consequence of a co-evolutionary arms race between the sexes, where males evolve more complex signals in response to females evolving resistance to those traits (Holland and Rice 1998). A female evolving under sexual antagonism could potentially mate with an allospecific male if their courtship display exploits female sensory biases or if the female is more likely to respond to novel signals from foreign males to whom she has not yet evolved resistance (Elias et al. 2012). Both of these scenarios could promote hybridization and would predict that the introgressing species with more complex courtship behaviour would be more likely to be the "donor". In fact, these predictions are consistent with previous observations of courtship and introgression in *Habronattus*. The "donor" species has a more complex courtship than the "recipient" species in instances of mitochondrial introgression detected in the *coecatus* group (Maddison and Hedin 2003) and the *amicus* group (Hedin and Lowder 2009).

Demographic characteristics of hybridizing species, such as differences in population sizes (Lepais et al. 2009) and dispersal behaviours (Funk and Omland 2003), could also cause asymmetric introgression. The abundant species is more likely to be the "donor" during nuclear introgression (Levin et al. 1996), which could explain the direction of nuclear introgression from *H. sansoni*/*H. americanus* to *H. ophrys*. Differences in population sizes can also have the reverse effect on mitochondrial introgression because females of the rarer species are more likely to mate with males of the abundant species, thus becoming the mitochondrial "donor" (Funk and Omland 2003). Female-biased dispersal could also produce a similar pattern (Wirtz 1999). Consequences of abundance differences between species on introgression directionality extend to biological invasions and range expansions (Currat et al. 2008). As a result, introgression patterns can provide insights on historical species distributions and changes. However, because most of our

introgression tests were not species-specific, inferences into historical events that caused introgression are difficult to make; thus, we can only describe the extent of historical introgression as opposed to particular events and movements of species in the past. More in-depth understanding about the natural history, relative species abundances, and geographic distributions of *Habronattus* would be useful to determine if demographic effects are major drivers of introgression. However, general trends of introgression observed here are consistent with limited information on the geographic proximity among introgressing species. In the VCC clade, for example, a lack of introgression involving the *viridipes* group could be attributed to their northern and eastern distribution. Most introgression that we detected in the VCC clade is constrained to species from the southern US and Mexico, where the *clypeatus* group and most of the *coecatus* group is distributed.

Many extant *Habronattus* live in sympatry (Griswold 1987), and these species could theoretically come into contact and hybridize. Species nearer to the Northern (e.g., Yukon) and Southern (e.g., Costa Rica) edges of the *Habronattus* range may be more geographically isolated, and have fewer opportunities for hybridization. On the other hand, species from the arid southwestern desert of the United States and northern Mexico, where *Habronattus* are most concentrated, usually have overlapping distributions with multiple other *Habronattus* species (Griswold 1987), or else are closely distributed so that a contact zone in the recent past could be inferred (e.g., *H. pugillis*, Maddison and McMahon 2000).

Divergent genitalia morphology may not prevent successful introgression between phylogenetically distant species. For example, our data indicates a history of hybridization between the *clypeatus* group and *coecatus* group, which possess very different genitalia.

Similarly, while reduced hybrid viability has been detected in divergent populations of *H. pugillis* (Masta and Maddison 2000), hybrids could be viable and backcross, resulting in distant introgression.

4.2.4. Evolutionary consequences of introgression

Introgression in *Habronattus* may have been frequent enough to result in detectable nuclear admixture, but infrequent enough so as not to behave as a homogenizing force reducing genetic and morphological diversity (Slatkin 1985). These findings agree with other studies in subpopulations of *H. americanus* (Maddison and Blackburn 2015) and *H. pugillis* (Maddison and McMahon 2000) showing strong selection for male ornaments and courtship differences in the face of gene flow. Distantly related species (e.g., the *clypeatus* and *coecatus* groups, *H.* sp. (ROBRT)), and closely related species with substantial admixture (e.g. *H. ophrys* and *H. americanus*/*H. sansoni)*, appear to have retained morphological distinctiveness despite a history of hybridization. Historical introgression in the species-rich *americanus* group (11 species) and the VCC group (38 species) does not appear to have slowed down diversification of these clades, both of which are composed of many species and have remarkable diversities of male ornaments.

While introgression may not have substantially reduced divergence among *Habronattus* species, it is unclear whether it has played a creative evolutionary role, either by promoting diversification or by influencing the phylogenetic distribution of traits. Distant introgression, which we detected in several clades, is more likely to have adaptive effects on lineages because novel and potentially adaptive genetic combinations are more likely to form as a result of introgression when there is more time to accumulate genetic differences (Stelkens and Seehausen 2009). Introgression could also create adaptive potential by increasing the standing variation of

hybridizing lineages, which could facilitate subsequent diversification (Seehausen 2013). There are an increasing number of documented cases of adaptive introgression (Huerta-Sánchez et al. 2014; Norris et al. 2015; Heliconius Consortium 2012) and introgression-facilitated diversification in animals (Streicher et al. 2014; Seehausen 2004). Given the strength of sexual selection in *Habronattus* (Masta and Maddison 2002), loci implicated in sexually selected traits could be under strong selection if they were exchanged between hybridizing species. There are several examples of morphological traits under sexual selection in *Habronattus* whose phylogenetic distributions could be explained by introgression (many are mentioned in sections 4.2.1 and 4.2.2). In particular, the red-purple bumps and dark spots only found on the 3rd legs of males of some populations of *H. sp.* (ROBRT) as well as in the *clypeatus* group member *H. velivolus* (with which *H.* sp. (ROBRT) is sympatric) and *H.* sp. (CHIH), and the distinctive "eyebrows" of *H. ophrys* and *H. sansoni* have phylogenetic distributions that match predictions based on introgression rather than convergence. What is most evident and interesting upon observation of these traits is the details of the trait shared amongst distantly related species. While almost all VCC group species have colourful ornaments on their 3rd legs (Griswold 1987), the similarity between the purple-red bumps and black spots found on the 3rd legs of *clypeatus* group species and *H.* sp. (ROBRT) is striking. It seems unlikely that this similarity evolved as a result of such refined convergence in female preferences. A possible explanation is that the trait was exchanged through introgression, and that the trait's details were retained in the recipient lineage. Hybridization has been previously invoked as the best explanation for homoplasy of male ornaments in hybridizing populations of *H.pugillis,* in part because homoplasy was found in the details of male ornaments between hybridizing populations (Maddison and McMahon 2000).

4.2.5. Future research

Sparse taxonomic sampling, both at the species and population level, limits our abilities to make inferences from D-statistics (Eaton et al. 2015). We did not include more than a single population per species and instead focused on constructing a more complete species tree. While we were able to produce a broad phylogeny of the group, all introgression signals are potentially affected by ghost taxa; therefore they cannot be attributed to specific species in most cases (Eaton et al. 2015). With a more complete phylogeny, the temporal and spatial sequence of introgression and speciation events can be more carefully mapped to the phylogeny, and historical biogeographical patterns and examples of adaptive introgression may emerge more readily (examples: Streicher et al. 2014; Eaton and Ree 2013; Cui et al. 2013). Increasing sampling at the population level would also test for effects of population choice (Pease and Hahn 2015; Eaton et al. 2015) and help identify biases arising from ancient population structure (Eriksson and Manica 2012) and population bottlenecks (Durand et al. 2011). This could be especially important for species of *Habronattus* with significant population structure (*pugillis* group: Maddison and McMahon 2000; *americanus* group: Maddison and Blackburn 2015). More in-depth sampling and testing for introgression between distant clades, particular between *H.* sp. (ROBRT) and the *clypeatus* group and between the *americanus* group and distant clades, could also help identify the timing of introgression relative to divergence of clades. This would clarify whether hybridization actually occurred between species with very different courtship behaviours, or if hybridization mostly occurred in ancestral lineages with less divergent courtship traits.

With introgression more fully characterized across *Habronattus*, future research should determine the extent of introgression's contribution to adaptation and diversification.

Widespread introgression in the highly diverse VCC clade, particularly between more distant groups *H. sp.* (ROBRT), *coecatus* and *clypeatus*, could be indicative of a correlation between introgression and diversification. Identifying the genetic basis of ornaments that may have introgressed is the next step to determine if hybridization played a role in their evolution. Because loci in this study have unknown genomic positions and low variability, it is unclear if the signal of introgression is spread throughout the genome, or if it is concentrated at particular regions. Understanding how the introgression signal is distributed and if there are signs of selection at introgressed loci would clarify whether introgression has been adaptive (Hedrick 2013).

5. Conclusions

We produced a highly resolved phylogeny for *Habronattus* and determined the relative contributions of hybridization and incomplete lineage sorting to genetic discordance in the group. We found that hybridization has been historically common in *Habronattus,* and has resulted in significant nuclear introgression in some instances (e.g., the *americanus* group, *H. sp.* (ROBRT)) and minimal nuclear introgression accompanied by strong mitochondrial introgression in others (e.g., the *coecatus* and *clypeatus* groups). Widespread introgression between both distant and closely related species indicates that only partial reproductive isolation has evolved across the *Habronattus* phylogeny. Frequent introgression may continue to occur between *Habronattus* species groups with divergent male ornaments and courtship behaviours (e.g. between the *clypeatus* group and the *coecatus* group), although ancestral introgression that precedes diversification of extant courtship behaviours and ornaments is also a possibility for some groups (e.g., *H.* sp. (ROBRT) and the *clypeatus* group, *americanus* group and distant clades).

Tables and figures

	Species 1	Species 2	Species 3	Species 4	Outgroup	Corresponding figure
(1) *americanus* group						
(1A)	*H. sansoni*	*H. americanus*	*H. ophrys*	*H. tarsalis*	*H. signatus*	11A
(1B)	*H. ophrys*	*H. tarsalis*	*H.* sp. (ESTU)	-	*H. signatus*	11B
(1C)	*H. sansoni*	*H. americanus*	*H.* sp. (ESTU)	-	*H. signatus*	11C
(2) *clypeatus* and *coecatus* group						
(2A)	*H. aztecanus*	*H. clypeatus*	*H. mexicanus*	*H. pyrrithrix*	*H. ophrys*	12A
(2B)	*H.* sp. (BLNDI)	*H. pyrrithrix*	*H. clypeatus*	*H. aztecanus*	*H. ophrys*	12B
(2C)	*H.* sp. (SLCTY)	*H. clypeatus*	*H.* sp. (BLNDI)	*H. pyrrithrix*	*H. ophrys*	12C
(2D)	*H.* sp. (SLCTY)	*H. clypeatus*	*H. mexicanus*	*H. pyrrithrix*	*H. ophrys*	12D
(2E)	*H.* sp. (SLCTY)	*H. clypeatus*	*H. borealis*	*H. pyrrithrix*	*H. ophrys*	12E
(2F)	*H. aztecanus*	*H. clypeatus*	*H. borealis*	*H. pyrrithrix*	*H. ophrys*	12F
(3) *H.* sp. (ROBRT) and VCC clade						
(3A)	*H.* sp. (SLCTY)	*H. jucundus*	*H.* sp. (ROBRT)	-	*H. ophrys*	13A
(3B)	*H. festus*	*H.* sp. (SLCTY)	*H.* sp. (ROBRT)	-	*H. ophrys*	13B
(3C)	*H. festus*	*H. jucundus*	*H.* sp. (ROBRT)	-	*H. ophrys*	13C
(4) H.decorus and *americanus* group						
(4A)	*H. ophrys*	*H.* sp. (ESTU)	*H. zapotecanus*	*H. decorus*	*H. signatus*	14A
(4B)	*H. ophrys*	*H.* sp. (ESTU)	*H. cambridgei*	*H. decorus*	*H. signatus*	14B
(4C)	*H. ophrys*	*H.* sp. (ESTU)	*H. oregonensis*	*H. decorus*	*H. signatus*	14C
(4D)	*H. ophrys*	*H.* sp. (ESTU)	*H. jucundus*	*H. decorus*	*H. signatus*	14D
(4E)	*H. ophrys*	*H.* sp. (ESTU)	*H. festus*	*H. decorus*	*H. signatus*	14E
(4F)	*H. zapotecanus*	*H. decorus*	*H. ophrys*	-	*H. signatus*	14A
(4G)	*H. cambridgei*	*H. decorus*	*H. ophrys*	-	*H. signatus*	14B
(4H)	*H. oregonensis*	*H. decorus*	*H. ophrys*	-	*H. signatus*	14C
(4I)	*H. jucundus*	*H. decorus*	*H. ophrys*	-	*H. signatus*	14D
(4J)	*H. festus*	*H. decorus*	*H. ophrys*	-	*H. signatus*	14E

Table 1 Species tested for introgression with Patterson's D statistic (4 species at a time) or DFOIL (5 species at a time). Results are shown in Figures 11-14.

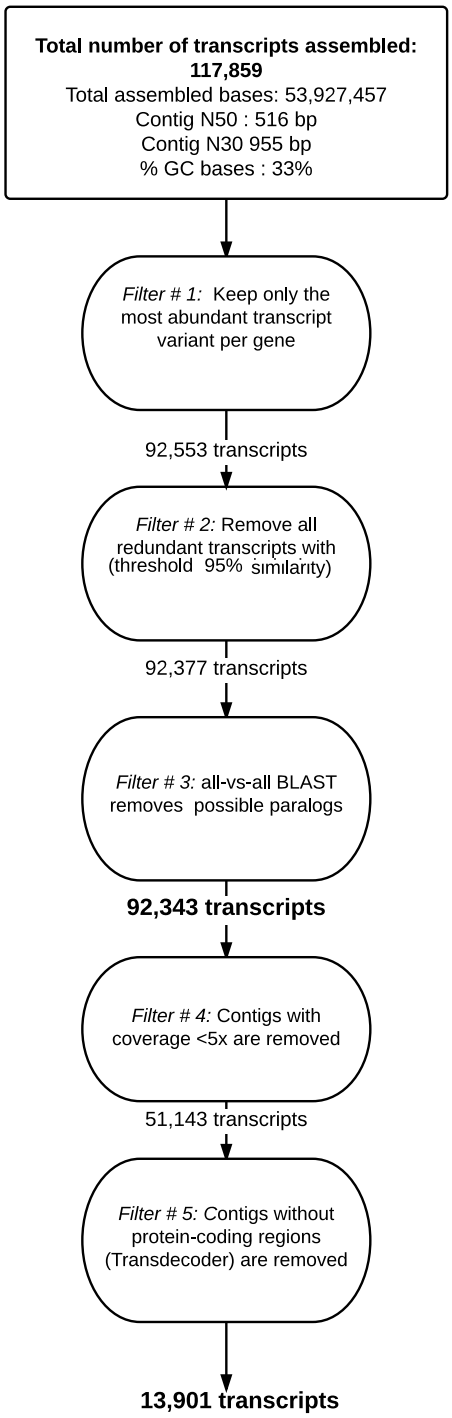


Figure 1. Summary of H. ophrys de novo transcriptome assembly and filtering steps (1. Abundance 2. Redundancy 3. Paralogy 4. Coverage 5. Coding Regions).

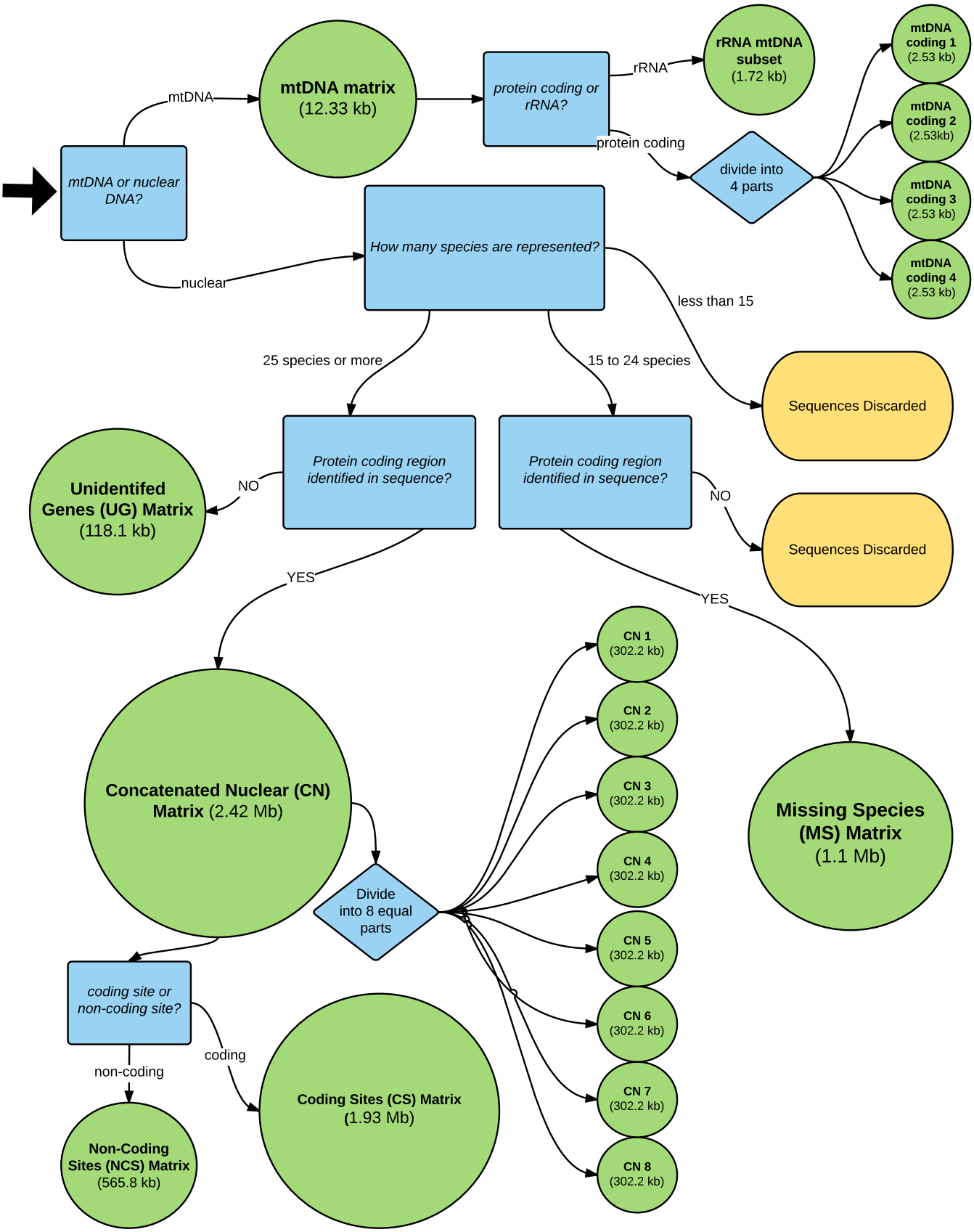


Figure 2. Flow chart describing how data was partitioned into matrices (green circles) for RAxML phylogenetic analyses.

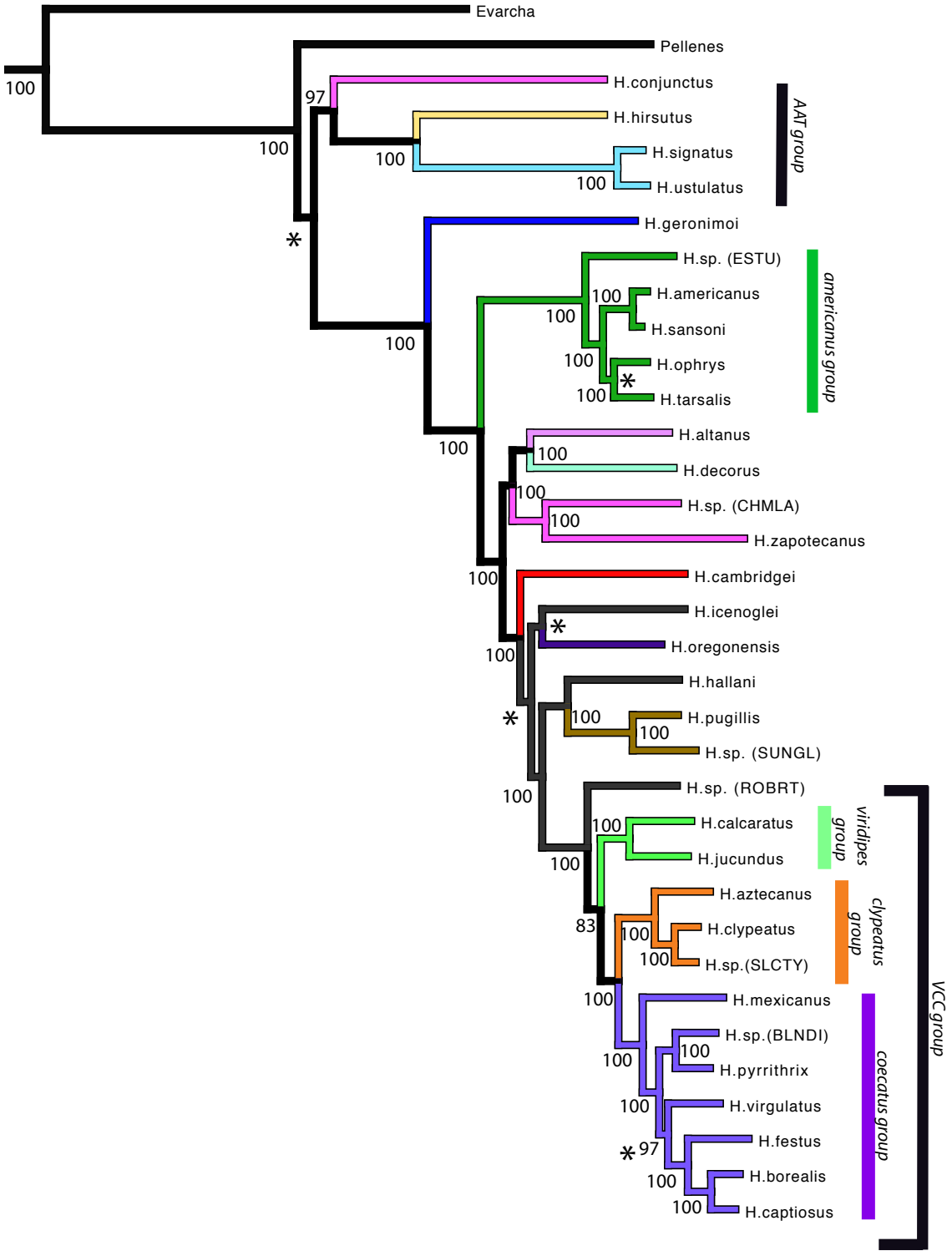
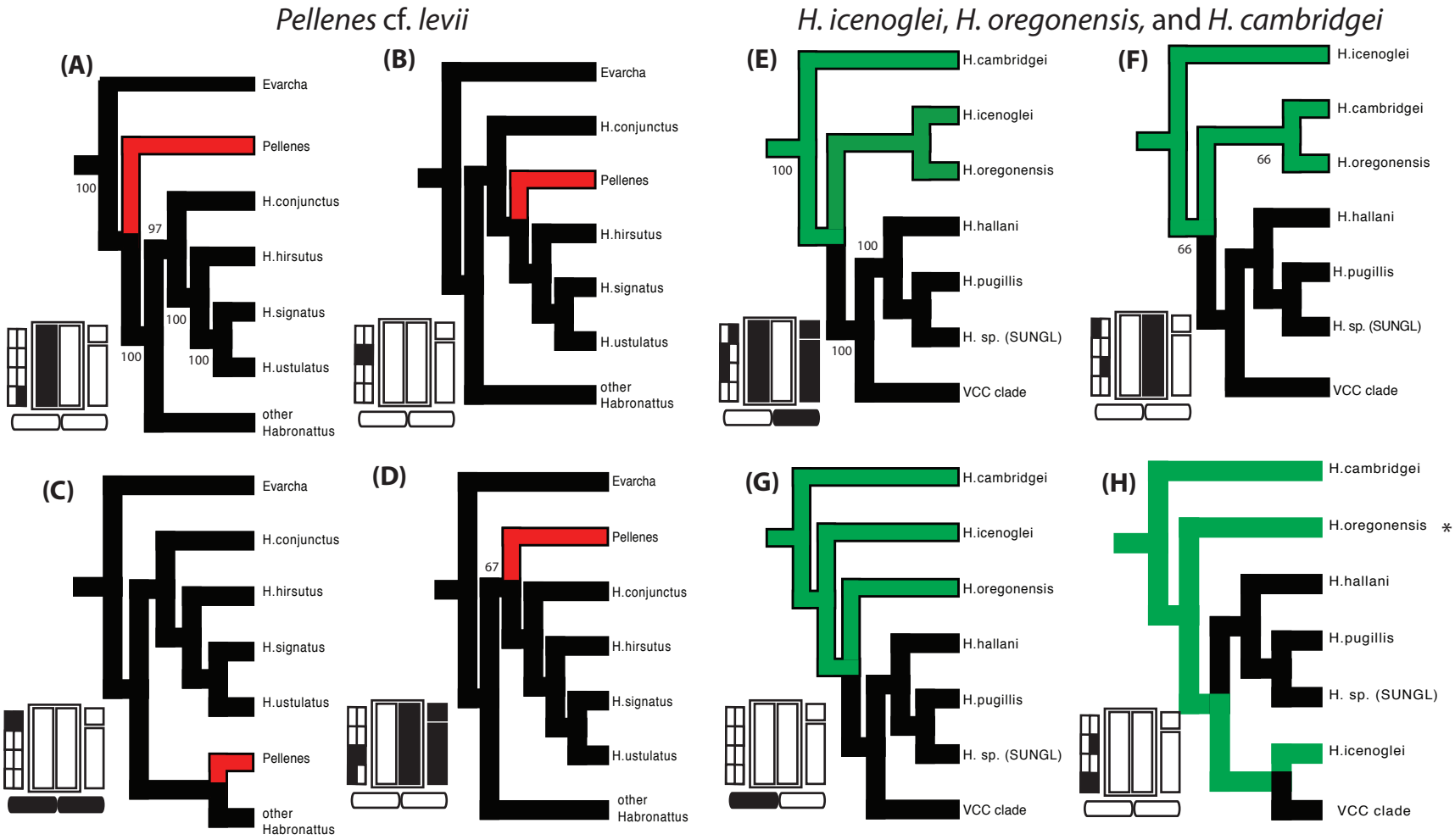


Figure 3. Concatenated nuclear Maximum likelihood phylogeny (lnL = -8016699). The matrix analyzed is a concatenated set of 1884 concatenated genes over 200 bp present in 25 or more species (total length = 2.41 Mbp). Branch lengths are proportional to change. Constructed in RAxML with 20 search replicates, 1000 bootstrap replicates, GTR+G+I substitution rate, partitioned as: codon pos1,pos2, pos3, noncoding. * indicate uncertain nodes that varied across nuclear ML analyses (i.e., phylogenies constructed from matrices composed of noncoding sites, coding sites, contigs present in 15-24 species, contigs without coding regions, subsets 1-8 matrices (described in Figure 4 and 5).



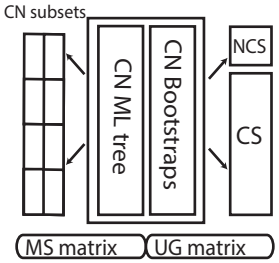


Figure 4. Common topologies at nodes that varied across maximum likelihood phylogenetic analyses. A dark square represents support from that particular matrix listed in the legend to the left: concatenated nuclear (CN) 3.18kb subsets, Concatenated Nuclear (CN) matrix ML tree and CN Bootstrap tree, Non-Coding Sites (NCS), Coding Sites (CS), Missing Species (15-24 species, MS), Unidentified Genes (no coding regions, UG). CN bootstrap values are at nodes. Phylogenies A-D show the position of Pellenes cf. levii relative to basal Habronattus species. Phylogenies E-G show the relationships of H. icenoglei and H. oregonensis to each other and to the "right" side of the phylogeny. The * by H. oregonensis at (H) indicates variable positions for this species amongst subsets.

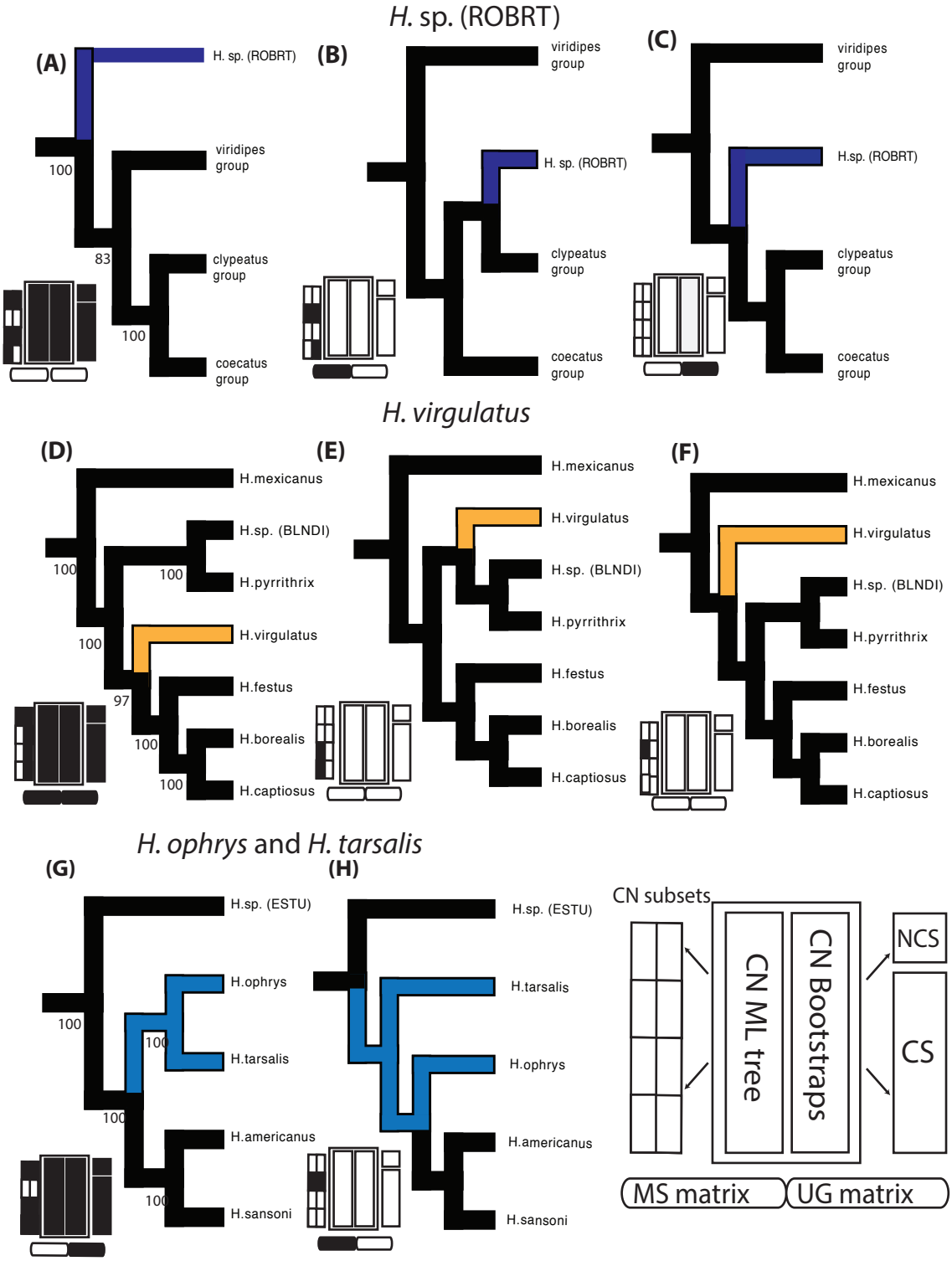


Figure 5. Common topologies at nodes that varied across maximum likelihood phylogenetic analyses. A dark square represents support from that particular matrix listed in the legend: Concatenated Nuclear (CN) 3.18 kb subsets, Concatenated Nuclear (CN) matrix ML tree, Non-Coding Sites (NCS), and Coding Sites (CS), Missing Species (MS), Unidentified genes (UG). CN bootstrap values are at nodes. Phylogenies A-C show the different placement of H. sp. (ROBRT) relative to the viridipes, clypeatus, and coecatus groups. Phylogenies D-F show the relationship of H. ophrys and H. tarsalis to the rest of the americanus group. Phylogenies G-H show the possible positions of H. virgulatus within the coecatus group.

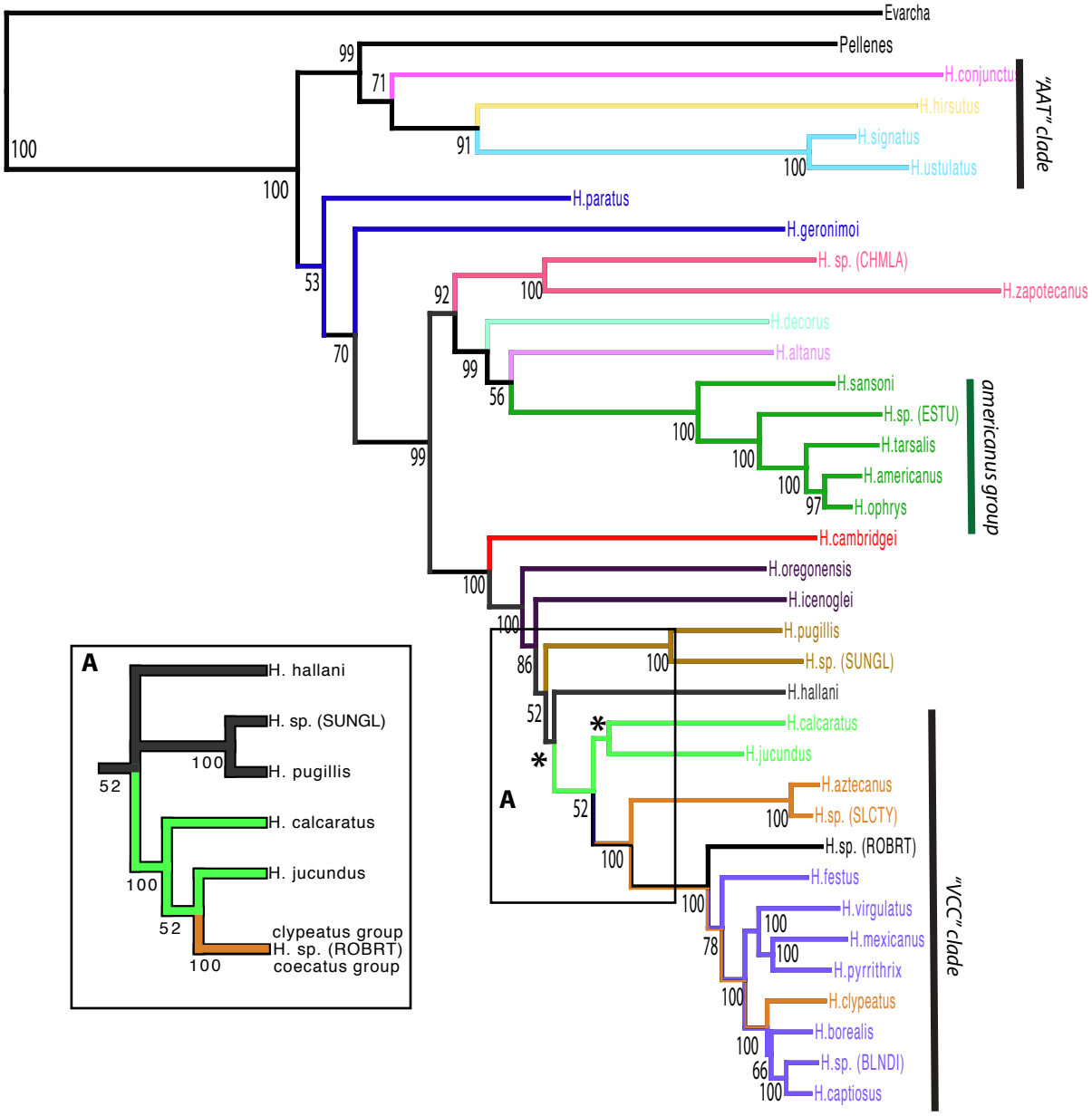


Figure 6. Concatenated Mitochondrial Maximum Likelihood phylogeny (12,233 bases, lnL = -3429108). Constructed in RAxML with 24 partitions (coding genes, rRNA, codon positions 1,2,3), 20 search replicates, 1000 bootstraps, GTR+G+I substitution rate. Branch lengths are proportional to change. Stars indicate nodes that were not recovered in the bootstrap Consensus tree (bootstrap consensus topology is shown in inset A).

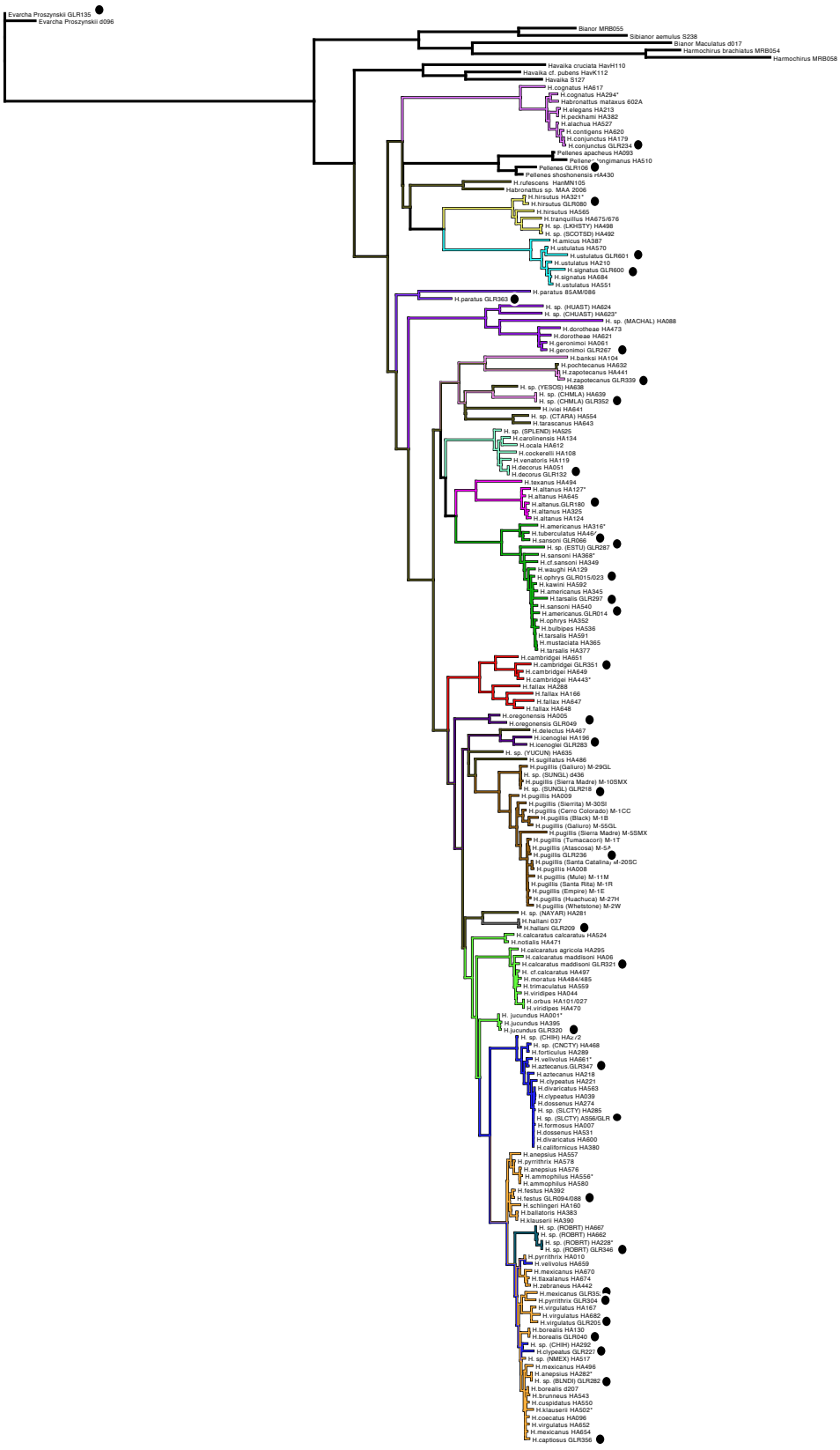


Figure 7. Maximum likelihood tree for the 16SND1 mitochondrial region (1047 bp). 160 Sanger sequenced specimens and all transcriptome specimens (marked by black circles) are included. The tree was constrained to the phylogenetic structure of the CM phylogeny (nodes with >90% bootstrap support only). Branch lengths are proportional to change. Constructed in RAxML with 20 search replicates, GTR+G+I substitution model. lnL = -25861.

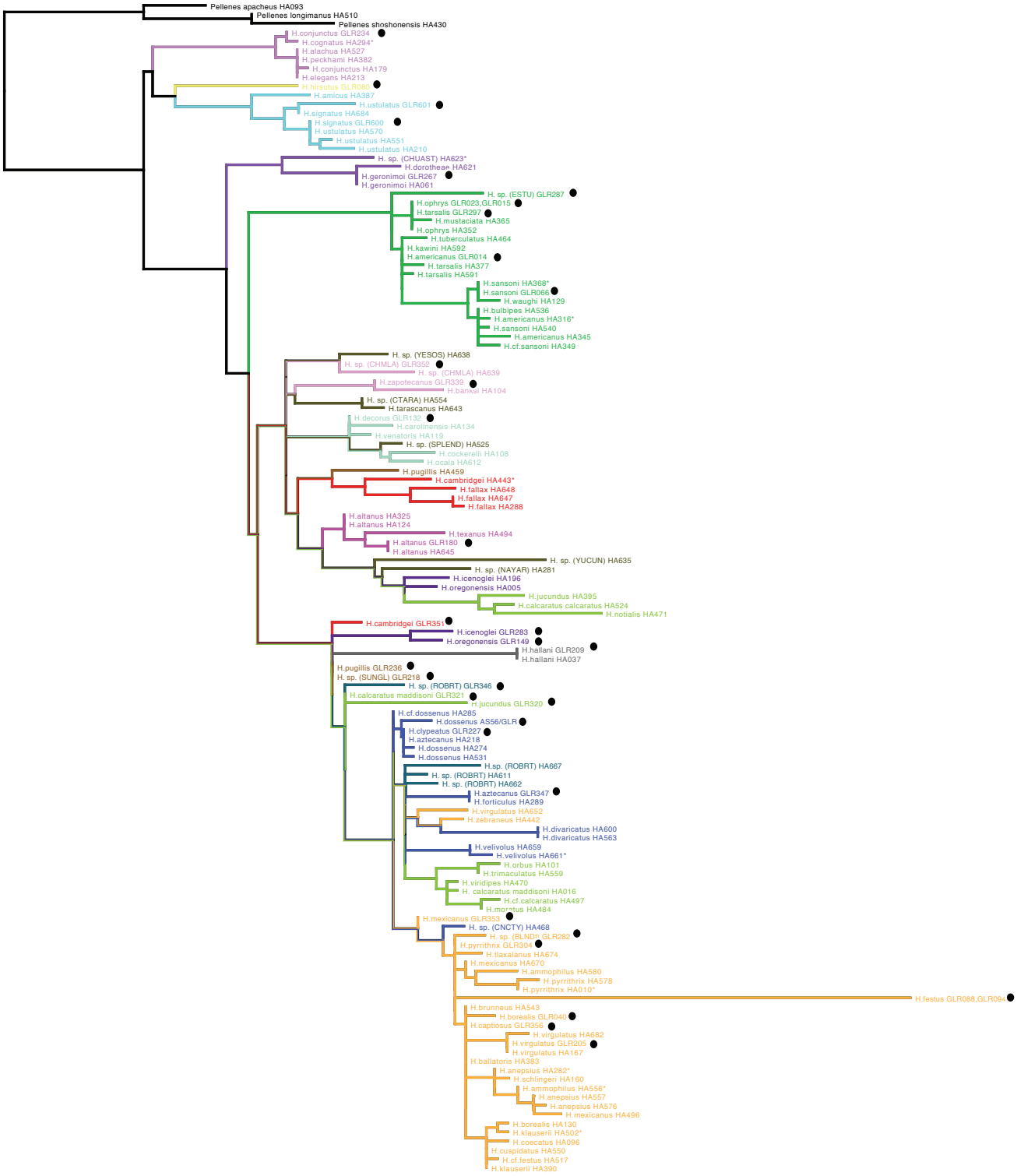


Figure 8. Maximum Likelihood tree for nuclear gene Ef1-a (819 bp). 105 Sanger sequenced specimens and all 33 Habronattus transcriptome specimens (marked by black circles) are included. Branch lengths are proportional to change. The tree was constrained to the phylogenetic structure of the concatenated nuclear phylogeny (nodes with >90% bootstrap support only). Constructed in RAxML with 20 search replicates and GTR+G+I substitution model, lnL = -25861.

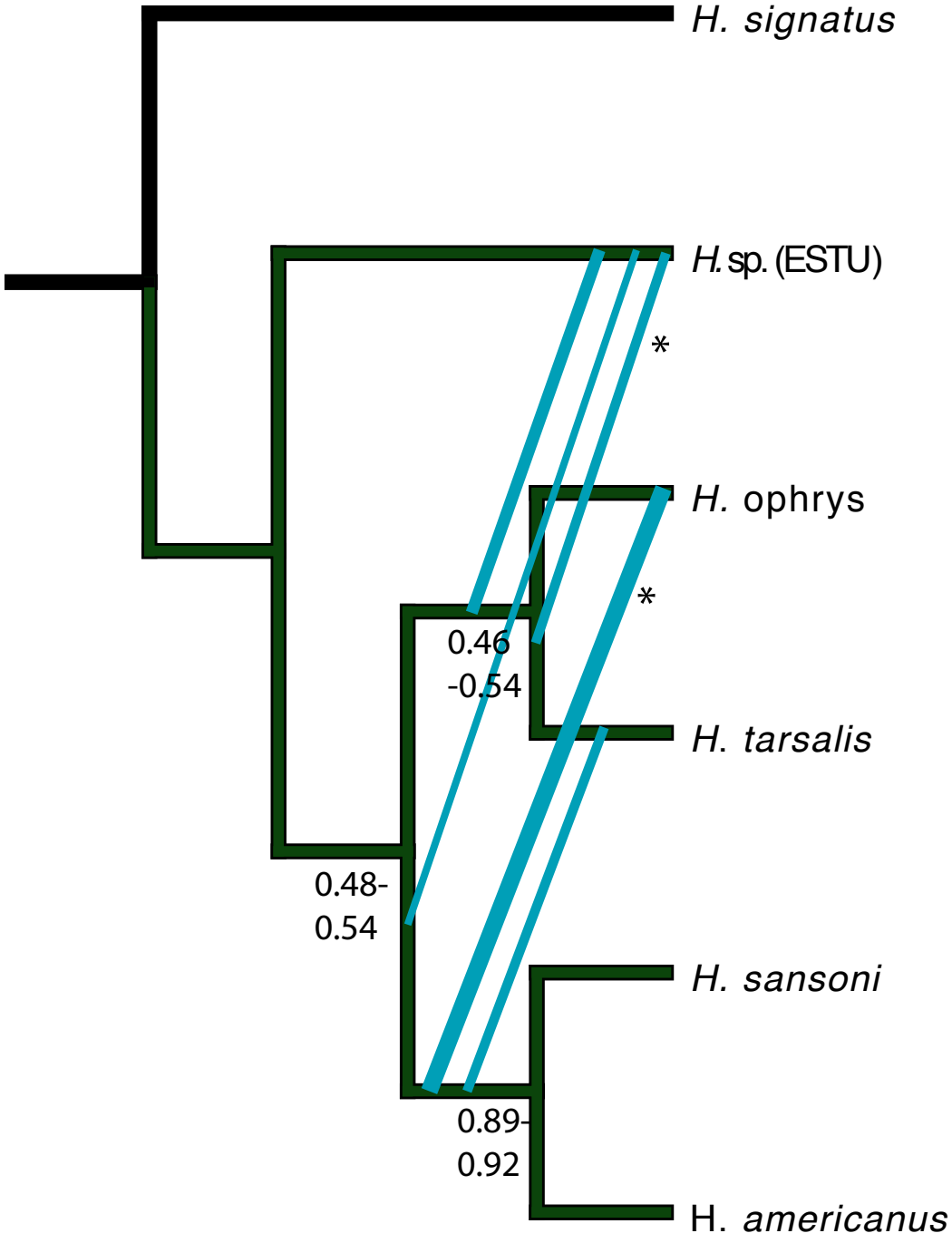


Figure 9. BUCKy Primary Concordance tree for the americanus group, based on 679 genes longer than 1 kb. Node values are the average estimated primary CFs for each clade. Diagonal lines represent non-dominant relationships with CFs credibility intervals overlapping 0.1. * indicates relationships also detected as introgression using D-statistics/DFOIL. The thickness of the diagonal line indicates the size of the CF (range from 0.084 to 0.305) and lines reaching an ancestral branch indicate introgression with all descendants. All CFs and their credibility intervals are listed in Appendix 3.

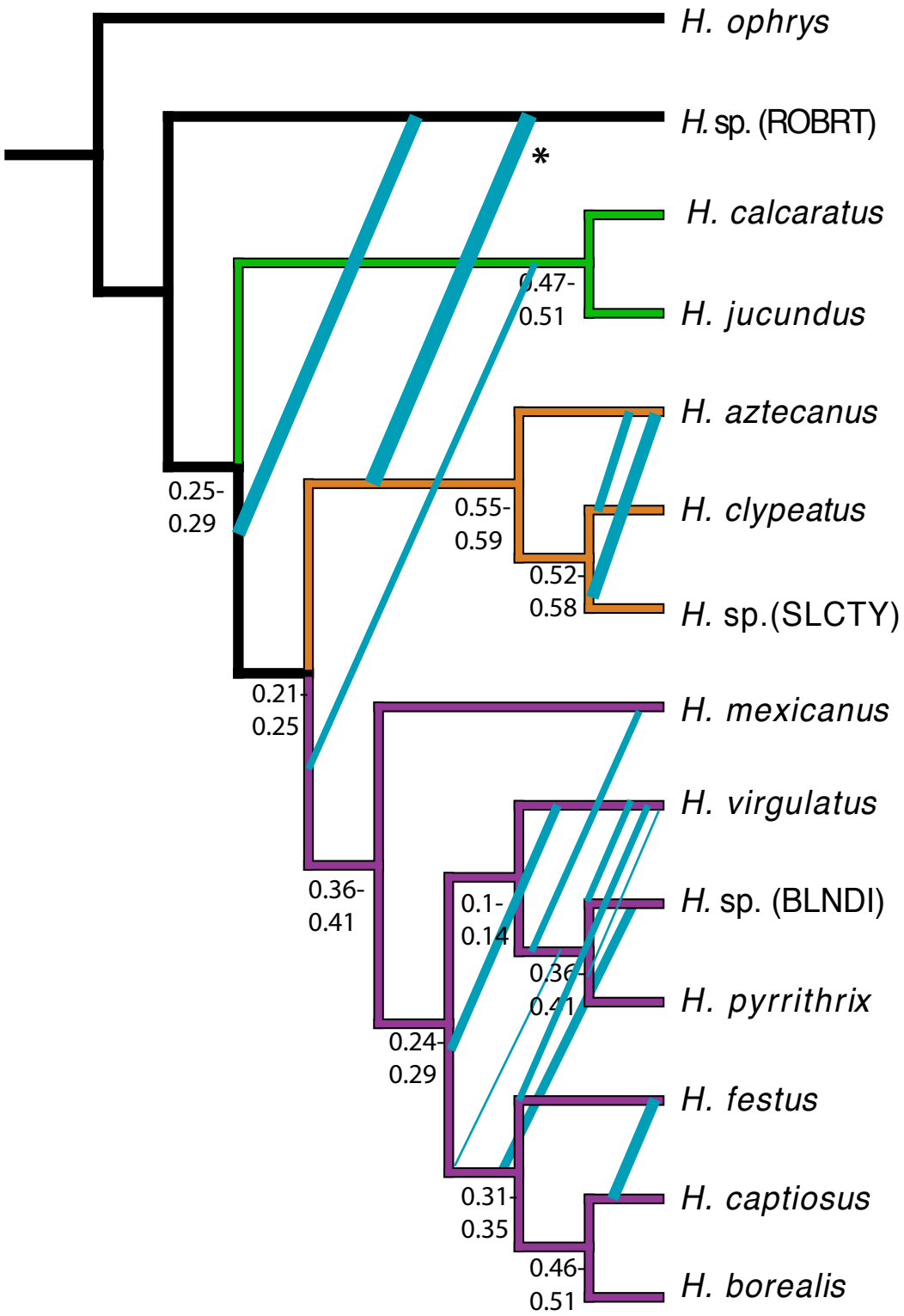


Figure 10. BUCKy Primary Concordance tree for the VCC clade, based on 517 genes longer than 1 kb. Node values are the CF credibility intervals for each clade. Diagonal lines represent non-dominant relationships with a CFs credibility intervals overlapping 0.1. The thickness of the diagonal line indicates the size of the CF (range from 0.196 to 0.084) and lines reaching an ancestral branch indicate introgression with all descendants. * indicates relationships detected as introgression using D-statistics/DFOIL (species tested listed in Table 1). All CFs and their credibility intervals are listed in Appendix 4.

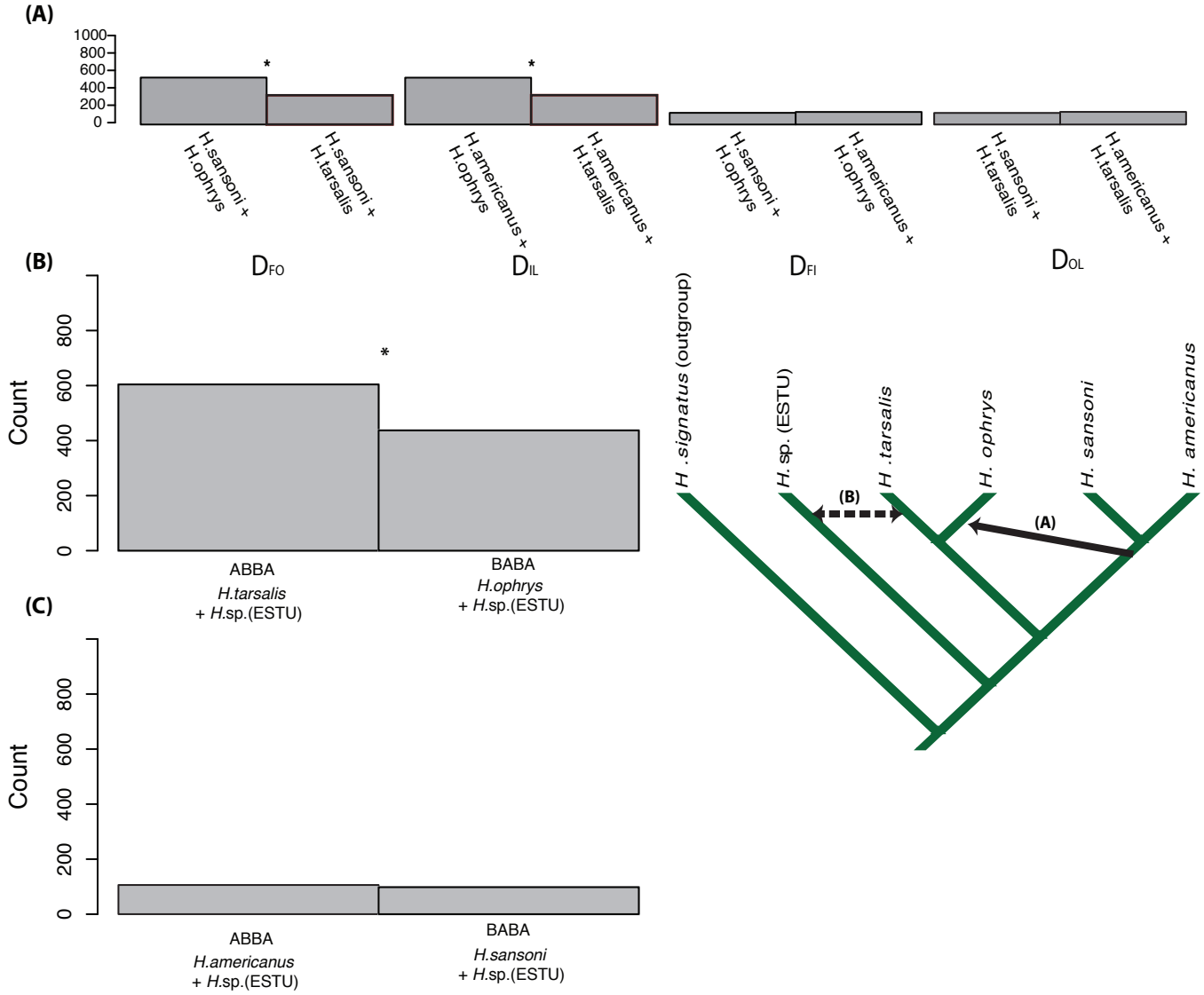


Figure 11. In (A), the total biallelic pattern counts for all DFOIL tests for introgression between americanus group species. In (B) and (C), ABBA vs. BABA allele patterns counts for D statistic tests for introgression americanus group species. Detected introgression is summarized on the phylogeny: dashed line represents support from D-statistics, solid line from DFOIL. The phylogenetic positions of each set of species are listed in Table 1.1A-C. p <0.00074 (indicated by *, 95% significance adjusted with Bonferroni correction for 68 tests) are from a chi-square binomial test and indicate that the left and right terms are significantly different. Allele counts are listed in Appendix 5 and 6, Dstatistics, DFO, DIL, DFI, DOL, and their p-values are listed in Appendix 7 and 8.

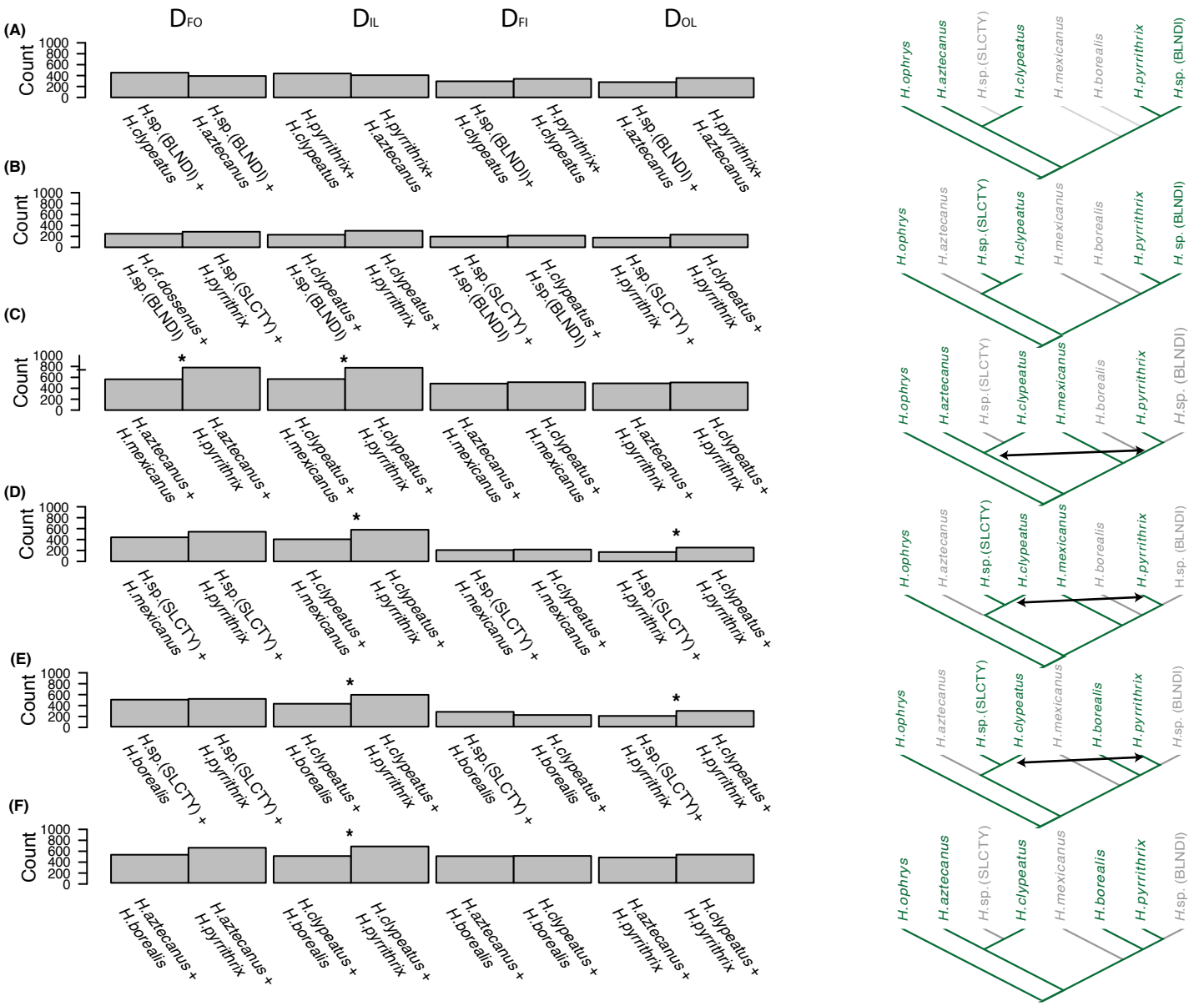


Figure 12. Total biallelic pattern counts for all DFOIL tests for introgression between the coecatus and clypeatus groups. p <0.00074 (indicated by *, 95% significance adjusted with Bonferroni correction for 68 tests) are from a chi-square binomial test and indicate that the left and right terms are significantly different. The phylogenetic positions of each set of species are listed in Table 1.2A-F. Detected introgression is summarized on the phylogeny (green branches represent the species being tested). Counts are listed in Appendix 5 and DFO, DIL, DFI, DOL, and their p-values are listed in Appendix 7.

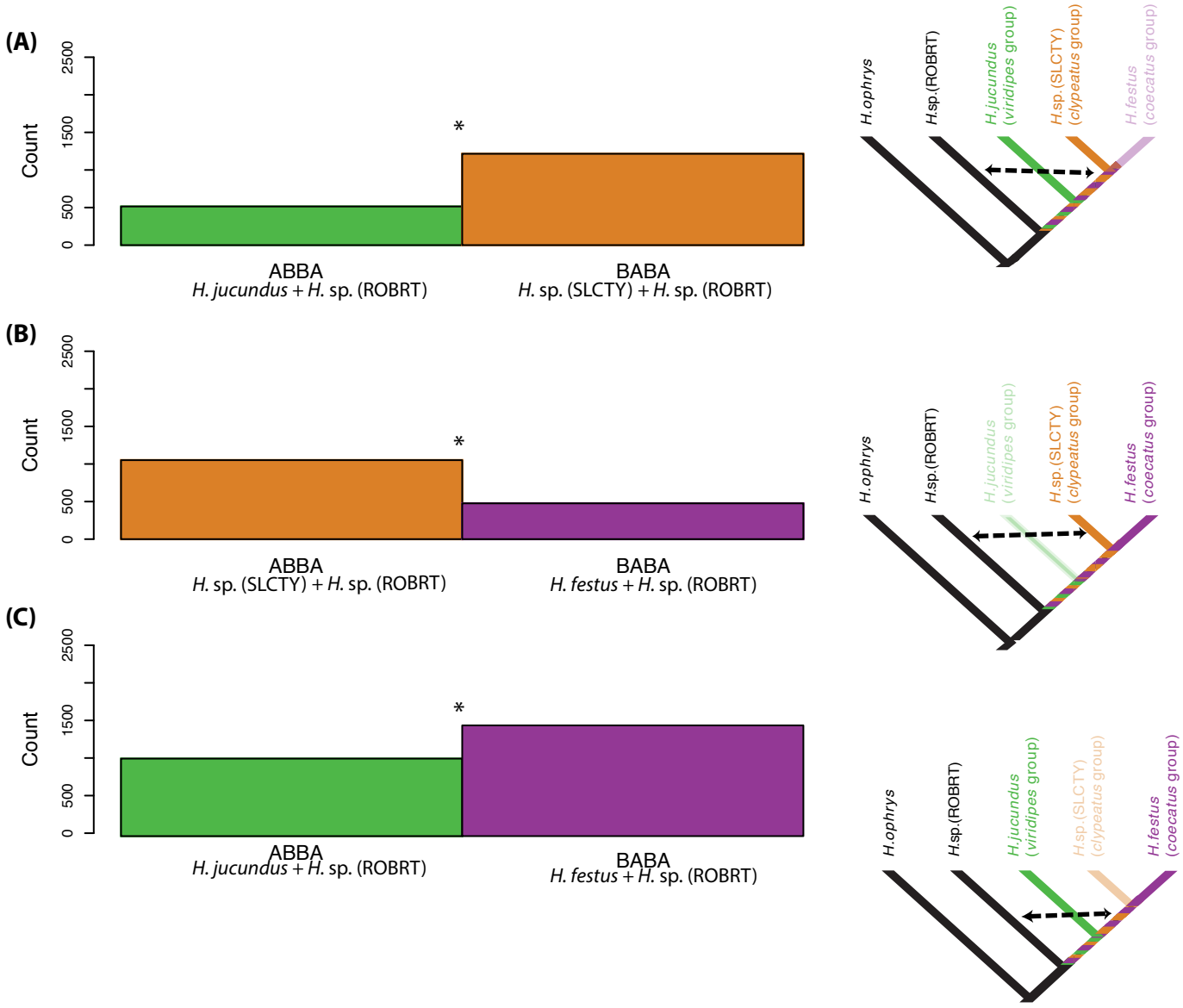


Figure 13. ABBA vs. BABA allele patterns counts for Partitioned D statistic tests for introgression between H. sp. (ROBRT) and the VCC clade. * indicate ABBA vs. BABA counts are significantly different (95% significance adjusted with Bonferroni correction for 68 tests to p <0.00074) from chi-square binomial test. Detected introgression is summarized on the phylogeny (dark branches represent the species being tested, light gray branches are excluded from the test). The phylogenetic positions of each set of species are listed in Table 1.1a-c. Counts are listed in Appendix 6, and D statistic and p values are listed in Appendix 8.

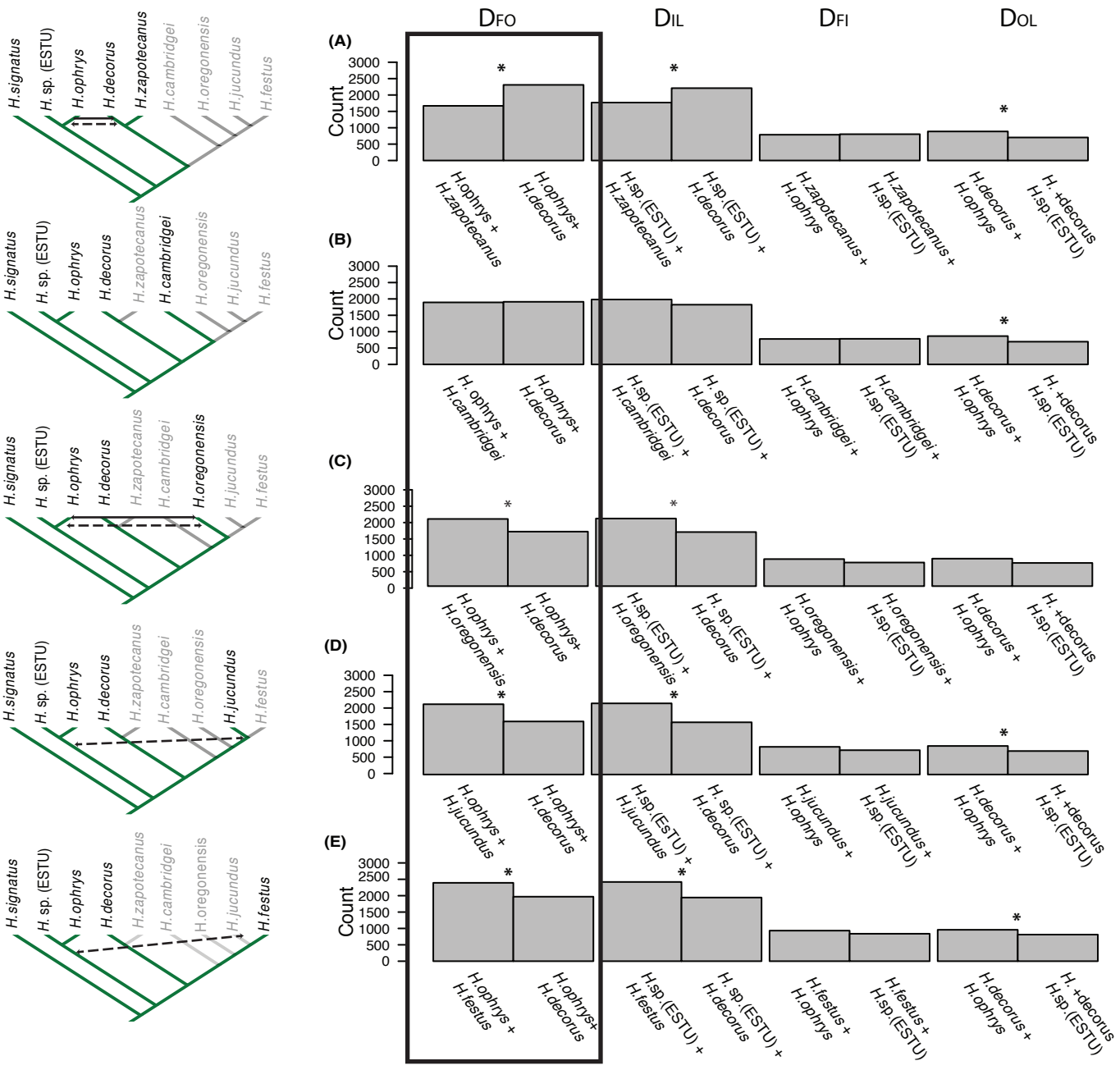


Figure 14. The total biallelic pattern counts for all DFOIL tests between americanus group, H. decorus, and species H. zapotecanus, H. cambridgei, H. oregonensis, H. jucundus, H. festus. Detected introgression is represented on the phylogeny: dashed line represents support from D-statistics, solid line from DFOIL. Green branches represent the species being tested. The phylogenetic positions of each set of species are listed in Table 1.4. p <0.00074 (indicated by *, 95% significance adjusted with Bonferroni correction for 68 tests) are from a chi-square binomial test and indicate that the left and right terms are significantly different. The black rectangle indicates the comparisonssupported by D-statistic tests (tests excluding H. sp. (ESTU)). Allele counts are listed in Appendix 5 and 6, D-statistics, DFO, DIL, DFI DOL, and their p-values are listed in Appendix 7 and 8.

Bibliography

Abbott, R., D. Albach, S. Ansell, J. W. Arntzen, S. J. E. Baird, N. Bierne, J. Boughman, A. Brelsford, C. A. Buerkle, R. Buggs, et al. 2013. Hybridization and speciation. J. Evol. Biol. 26:229–246.

Anderson, E. 1949. Introgressive hybridization. John Wiley and Sons, Inc.

Andrews, S. 2010. FastQC: A quality control tool for high throughput sequence data.; http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Ané, C., B. Larget, D. Baum, S. D. Smith, and A. Rokas. 2007. Bayesian estimation of concordance among gene trees. Mol. Biol. Evol. 24:412–26.

Ané, C. 2011. Reconstructing concordance trees and testing the coalescent model. In Knowles, L. L., & Kubatko, L. S. (Eds.). Estimating species trees: practical and theoretical aspects. John Wiley and Sons.

Aronesty, E. 2011. ea-utils: Command-line tools for processing biological sequencing data.; http://code.google.com/p/ea-utils

Avise, J. 2001. Cytonuclear genetic signatures of hybridization phenomena: rationale, utility, and empirical examples from fishes and other aquatic animals. Rev. Fish Biol. Fish. 10:253–263.

Baack, E. J., and L. H. Rieseberg. 2007. A genomic view of introgression and hybrid speciation. Curr. Opin. Genet. Dev. 17:513–8.

Baum, D. A. 2007. Concordance trees , concordance factors , and the exploration of reticulate genealogy. 56:417–426.

Blackburn, G. S., and W. P. Maddison. 2014. Stark sexual display divergence among jumping spider populations in the face of gene flow. Mol. Ecol. 23:5208–5223.

Boeckmann B., A. Bairoch, R. Apweiler, M. Blatter, A. Estreicher, E. Gasteiger, M. J. Martin, K. Michoud, C. O'Donovan, I. Phan, et al. 2003. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res. 31:365-370.

Buchholz, S. 2007. A First Contribution to the Arachnofauna (Arachnida: Araneae) of the Nestos Delta (NE Greece). Acta Zool. Bulg. 59:241–252.

Chung, Y., and Ané, C. 2011. Comparing two Bayesian methods for gene tree/species tree reconstruction: simulations with incomplete lineage sorting and horizontal gene transfer. Systematic Biology, 60:261-275.

Cui, R., M. Schumer, K. Kruesi, R. Walter, P. Andolfatto, and G. G. Rosenthal, 2013. Phylogenomics reveals extensive reticulate evolution in Xiphophorus fishes. Evolution, 67:2166-2179

Currat, M., M. Ruedi, R. J. Petit, and L. Excoffier. 2008. The hidden side of invasions: massive introgression by local genes. Evolution 62:1908–20.

Degnan, J., and N. Rosenberg. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24:332–340.

Donnelly, M. J., J. Pinto, R. Girod, N. J. Besansky, and T. Lehmann. 2004. Revisiting the role of introgression vs shared ancestral polymorphisms as key processes shaping genetic diversity in the recently separated sibling species of the Anopheles gambiae complex. Heredity. 92:61–68.

Dunn, C. W., A. Hejnol, D. Q. Matus, K. Pang, W. E. Browne, S. A Smith, E. Seaver, G. W. Rouse, M. Obst, G. D. Edgecombe, et al. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452:745–749.

Durand, E., N. Patterson, D, Reich, and M. Slatkin. 2011. Testing for ancient admixture between closely related populations. Mol. Biol. Evol. 28:2239–2252.

Eaton, D. A. R., and R. H. Ree. 2013. Inferring phylogeny and introgression using RADseq data: an example from flowering plants (Pedicularis: Orobanchaceae). Syst. Biol. 62:689–706.

Eaton, D., A. Hipp, A. González-Rodríguez, and J. Cavender-Bares. 2015. Historical introgression among the American live oaks and the comparative nature of tests for introgression. Evolution. 69:2587–2601.

Elias, D. O. 2006. Female preference for complex/novel signals in a spider. Behav. Ecol. 17:765–771.

Elias, D. O., W. P. Maddison, C. Peckmezian, M. B. Girard, and A. C. Mason. 2012. Orchestrating the score: complex multimodal courtship in the Habronattus coecatus group of Habronattus jumping spiders (Araneae: Salticidae). Biol. J. Linn. Soc. 105:522–547.

Elias, D., A. Mason, and W. Maddison. 2003. Seismic signals in a courting male jumping spider (Araneae: Salticidae). J. Exp. Biol. 206:4029–4039.

Eriksson, A., and A. Manica. 2012. Effect of ancient population structure on the degree of polymorphism shared between modern human populations and ancient hominins. Proc. Natl. Acad. Sci. U. S. A. 109:13956–13960.

Fiser, C., and G. N. Azarkina. 2005. A Contribution to the Knowledge of the Jumping Spiders (Salticidae: Araneae) of the Republic of Macedonia. Acta Zool. Bulg. 57:299–304.

Fu, L., B. Niu, Z. Zhu, S. Wu, and W. Li. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150–3152.

Funk, D. J., and K. E. Omland. 2003. Species-Level Paraphyly and Polyphyly: Frequency, Causes, and Consequences, with Insights from Animal Mitochondrial DNA. Annu. Rev. Ecol. Evol. Syst. 34:397–423.

Good, J., D. Vanderpool, S. Keeble, and K. Bi. 2015. Negligible nuclear introgression despite complete mitochondrial capture between two species of chipmunks. Evolution. 69:1961–1972.

Grant, P. R., B. R. Grant, and K. Petren. 2005. Hybridization in the recent past. Am. Nat. 166:56–67.

Gregory, T. R., and D. P. Shorthouse. 2003. Genome Sizes of Spiders. J. Hered. 94:285–290.

Griswold, C. 1987. A revision of the jumping spider genus Habronattus FOP-Cambridge (Araneae; Salticidae), with phenetic and cladistic analyses. Univ. Calif. Publ. Entomol. 107:1–344.

Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li, M. Lieber, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8:1494–512.

Hebets, E., and W. P. Maddison. 2005. Xenophilic mating preferences among populations of the jumping spider Habronattus pugillis Griswold. Behav. Ecol. 16:981–988.

Hedin, M., and M. Lowder. 2009. Phylogeography of the Habronattus amicus species complex (Araneae: Salticidae) of western North America, with evidence for localized asymmetrical mitochondrial. Zootaxa 2307:39–60.

Hedin, M. C., and W. P. Maddison. 2000. Phylogenetic Utility and Evidence for Multiple Copies of Elongation Factor-1α in the Spider Genus Habronattus (Araneae : Salticidae). 18:1512–1521.

Hedrick, P. 2013. Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. Mol. Ecol. 22:4606–4618.

Heliconius Genome Consortium. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature. 487:94-98.

Holland, B., and W. Rice. 1998. Perspective: chase-away sexual selection: antagonistic seduction versus resistance. Evolution. 52:1–7.

Huerta-Sánchez H., X. Jin, A. and Z. Bianba, B.M. Peter, N. Vinckenbosch, Y. Liang, X. Yi, M. He, M. Somel, P. Ni, et al. 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature. 512:194-197.

Joly, S., P. A. McLenachan, and P. J. Lockhart. 2009. A statistical approach for distinguishing hybridization and incomplete lineage sorting. Am. Nat. 174:54–70.

Katoh, K., and D. Standley. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30:3059–3066.

Kronforst, M., L. Young, L. Blume, and L. Gilbert. 2006. Multilocus analyses of admixture and introgression among hybridizing Heliconius butterflies. Evolution. 60:1254–1268.

Kubatko, L. S., and J. H. Degnan. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56:17–24.

Lanfear, R., B. Calcott, S. Ho, and S. Guindon. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29:1695–1701.

Larget, B., S. Kotha, C. Dewey, and C. Ané. 2010. BUCKy: gene tree/species tree reconciliation with Bayesian concordance analysis. Bioinformatics 26:2910–2911.

Lemmon, A. R., and E. M. Lemmon. 2012. High-throughput identification of informative nuclear loci for shallow-scale phylogenetics and phylogeography. Syst. Biol. 61:745–61.

Lepais, O., R. J. Petit, E. Guichoux, J. E. Lavabre, F. Alberto, A. Kremer, and S. Gerber. 2009. Species relative abundance and direction of introgression in oaks. Mol. Ecol. 18:2228–42.

Levin, D. A., J. Francisco-Ortega, and R. K. Jansen. 1996. Hybridization and the Extinction of Rare Plant Species. Conserv. Biol. 10:10–16.

Li, B., and C. Dewey. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323.

Logunov, D. 1999. A review of the genus Pellenes in the fauna of Central Asia and the Caucasus (Araneae, Salticidae). J. Nat. Hist. 33:89–148.

Maddison, W. P. 1982. XXXY sex chromosomes in males of the jumping spider genus Pellenes (Araneae: Salticidae). Chromosoma 85:23–37.

Maddison, W. P. 1997. Gene trees in species trees. Systematic biology, 46:523-536.

Maddison, W. P., and G. S. Blackburn. 2015. Insights to the mating strategies of Habronattus americanus jumping spiders from natural behaviour and staged interactions in the wild. Behaviour 152:1169–1186.

Maddison, W. P., and M. Hedin. 2003. Phylogeny of Habronattus jumping spiders (Araneae : Salticidae), with consideration of genitalic and courtship evolution. Syst. Entomol. 28:1–22.

Maddison, W. P., and G. Leduc-Robert. 2013. Multiple origins of sex chromosome fusions correlated with chiasma localization in Habronattus jumping spiders (Araneae: Salticidae). Evolution 67:2258–72.

Maddison, W. P., and D. Maddison. 2015. Mesquite: a modular system for evolutionary analysis. Version 2.75. 2011.

Maddison, W. P., and M. McMahon. 2000. Divergence and Reticulation among Montane Populations of a Jumping Spider (Habronattus pugillis Griswold). Syst. Biol. 49:400–421.

Mallet, J. 2005. Hybridization as an invasion of the genome. Trends Ecol. Evol. 20:229–37.

Martin, S., J. Davey, and C. Jiggins. 2015. Evaluating the use of ABBA-BABA statistics to locate introgressed loci. Mol. Biol. Evol. 32:244–257.

Masta, S. E., and J. L. Boore. 2004. The complete mitochondrial genome sequence of the spider Habronattus oregonensis reveals rearranged and extremely truncated tRNAs. Mol. Biol. Evol. 21:893–902.

Masta, S. E., and W. P. Maddison. 2002. Sexual selection driving diversification in jumping spiders. Proc. Natl. Acad. Sci. U.S.A. 99:4442–4447.

Miller, J. R., S. Koren, and G. Sutton. 2010. Assembly algorithms for next-generation sequencing data. Genomics 95:315–327.

Norris, L.C., B.J. Main, Y. Lee, T.C. Collier, A. Fofana, A.J. Cornel, and G.C. Lanzaro. 2015. Adaptive introgression in an African malaria mosquito coincident with the increased usage of insecticide-treated bed nets. Proc. Natl. Acad. Sci. USA. 112:815-820.

Pamilo, P., and M., Nei. 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5:568-583.

Panhuis, T. M., R. Butlin, M. Zuk, and T. Tregenza. 2001. Sexual selection and speciation. Trends Ecol. Evol. 16:364–371.

Pease, J., and M. Hahn. 2015. Detection and Polarization of Introgression in a Five-taxon Phylogeny. Syst. Biol. 64:651–662.

Peckham, G., and E. Peckham. 1889. Observations on sexual selection in spiders of the family Attidae. Ornamentation. Occas. Pap. Nat. Hist. Soc. Wisconsin. 1:117-151.

Peckham, G., and E. Peckham. 1890. Additional Observations on Sexual Selection in Spiders of the Family Attidae: With Some Remarks on Mr. Wallace's Theory of Sexual Ornamentation. Occas. Pap. Nat. Hist. Soc. Wisconsin. 1:3-60.

Rieseberg, L., O. Raymond, D. Rosenthal, and Z. Lai. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. Science. 301:1211–1216.

Ronquist, F., and J. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.

Ronquist, F., J. Huelsenbeck, and P. van der Mark. 2005. MrBayes 3.1 Manual.

Sanggaard, K.W., J.S Bechsgaard, X. Fang, J. Duan, T.F. Dyrlund, V. Gupta, X. Jiang, L. Cheng, D. Fan, Y. Feng, et al. 2014. Spider genomes provide insight into composition and evolution of venom and silk. Nature communications 5:3765.

Sankararaman, S., N. Patterson, H. Li, S. Pääbo, and D. Reich. 2012. The date of interbreeding between Neandertals and modern humans. PLoS Genet. 8:e1002947.

Scheidemantel, D. D. 1997. Behavioral and natural history studies of the jumping spider Habronattus oregonensis and inquiry based secondary laboratory lesson development stemming from university research. PhD. Diss., The University of Arizona, Arizona.

Schluter, D., and T. Price. 1993. Honesty, perception and population divergence in sexually selected traits. Proc. R. Soc. London B Biol. Sci. 253:117–122.

Schmieder, R., and R. Edwards. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27:863–4.

Seehausen, O. 2004. Hybridization and adaptive radiation. Trends Ecol. Evol. 19:198–207.

Seehausen, O. 2013. Conditions when hybridization might predispose populations for adaptive radiation. J. Evol. Biol. 26:279–81.

Slatkin, M. 1985. Gene flow in natural populations. Annu. Rev. Ecol. Syst. 393–430.

Smith, S., M. Moore, and J. Brown. 2015. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. BMC Evol. Biol. 15:150.

Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.

Stelkens, R., and O. Seehausen. 2009. Genetic distance between species predicts novel trait expression in their hybrids. Evolution 63:884–97.

Streicher, J. W., T. J. Devitt, C. S. Goldberg, J. H. Malone, H. Blackmon, and M. K. Fujita. 2014. Diversification and asymmetrical gene flow across time and space: lineage sorting and hybridization in polytypic barking frogs. Mol. Ecol. 23:3273–3291.

Toews, D. P. L., and A. Brelsford. 2012. The biogeography of mitochondrial and nuclear discordance in animals. Mol. Ecol. 21:3907–30.

Turelli, M., N. H. Barton, and J. A. Coyne. 2001. Theory and speciation. Trends Ecol. Evol. 16:330–343.

Twyford, A. D., and R. A. Ennos. 2012. Next-generation hybridization and introgression. Heredity (Edinb). 108:179–189.

Wirtz, P. 1999. Mother species-father species: unidirectional hybridization in animals with female choice. Anim. Behav. 58:1–12.

Yu, Y., C. Than, J. H. Degnan, and L. Nakhleh. 2011. Coalescent histories on phylogenetic networks and detection of hybridization despite incomplete lineage sorting. Syst. Biol. 60:138–49.

Zink, R. M., & Barrowclough, G. F. 2008. Mitochondrial DNA under siege in avian phylogeography. Molecular ecology, 17:2107-2121.

Zurek, D. B., T. W. Cronin, L. A. Taylor, K. Byrne, M. L. G. Sullivan, and N. I. Morehouse. 2015. Spectral filtering enables trichromatic vision in colorful jumping spiders. Curr. Biol. 25:R403–R404.

Appendices

Appendix 1 Specimens from which transcriptomes were obtained, their localities, and their sex chromsomes (if they are known. Only mitochondrial data was gathered for *H.paratus*.

Species	Voucher #	Sex	Locality	Latitude , Longitude	Species karyotype (male sex chromosomes only)
coecatus group					
H. festus	GLR094	M	Hayne's Lease Ecological Reserve, near Osoyoos, BRITISH COLUMBIA, CANADA	49.0813, -119.5181	XXO
H. festus	GLR088	F	Hayne's Lease Ecological Reserve, near Osoyoos, BRITISH COLUMBIA, CANADA	49.0813, -119.5181	XXO
H. virgulatus	GLR205	M	Mt. Hopkins Road, Santa Cruz Co., ARIZONA, U.S.A.	31.689, -110.975	XXO
H. borealis	GLR040	M	Hamilton Beach Strip, Burlington, ONTARIO, CANADA	43.33, -79.8	XXXY
H. mexicanus	GLR353	M	Rancho Primavera, El Tuito, JALISCO, MEXICO	20.341, -105.350	XXY
H. pyrrithrix	GLR304	M	Levee Road, Sunrise Point Park, Yuma County., ARIZONA, U.S.A.	32.731, -114.612	XXO
H. captiosus	GLR356	F	shore of Little Smoky River at HWY 49, S of Guy, ALBERTA, CANADA	55.4505, -117.1440	XXXY

Species	Voucher #	Sex	Locality	Latitude , Longitude	Species karyotype (male sex chromosomes only)
H. sp. (BLNDI)	GLR282	M	Estero Morúa, Puerto Peñasco, SONORA, MEXICO	31.293, -113.452	uncertain
americanus group					
H. ophrys	GLR023	M	Iona Beach, Richmond, BRITISH COLUMBIA, CANADA	49.221, -123.214	XXO
H. ophrys	GLR015	F	Iona Beach, Richmond, BRITISH COLUMBIA, CANADA	49.221, -123.214	XXO
H. americanus	GLR014	M	Iona Beach, Richmond, BRITISH COLUMBIA, CANADA	49.221, -123.214	XXO
H. tarsalis	GLR297	M	Levee Road, Sunrise Point Park,Yuma Co., ARIZONA, U.S.A.	32.731, -114.612	XXO
H. sansoni	GLR066	M	Borgata Lodge, Kelowna	49.954, -119.398	XXO
H. sp. (ESTU)	GLR287	F	Estero Cerro Prieto, SONORA, MEXICO	31.418, -113.626	uncertain
clypeatus group					
H. clypeatus	GLR227	M	Mt. Hopkins Road, Santa Cruz Co., ARIZONA, U.S.A.	31.686, -110.975	uncertain
H. aztecanus	GLR347	M	Bocanegra Beach, Puerto Vallarta, JALISCO, MEXICO	20.670, -105.274	XXO
H. sp. (SLCTY)	AS56	F	Silver City, NEW MEXICO, U.S.A.	n/a	uncertain

Species	Voucher #	Sex	Locality	Latitude , Longitude	Species karyotype (male sex chromosomes only)
amicus group					
H. signatus	GLR600	n/a	Ocotillo, CALIFORNIA, U.S.A.	32.7421, -115.9949	XXO
H. ustulatus	GLR601	n/a	Cleveland National Forest, CALIFORNIA, U.S.A.	32.7302, -116.4607	uncertain
viridipes group					
H. jucundus	GLR320	M	Bolan lake, Joseph County, OREGON, U.S.A.	42.024, -123.461	XXO
H. calcaratus maddisoni	GLR321	M	West Road near HWY 11, Hailey Byw W. , Haileybury, ONTARIO, CANADA	47.45, -79.708	XXXY
oregonensis group					
H. oregonensis	GLR149	M	Tantalus Lookout Road, Squamish, BRITISH COLUMBIA, CANADA	49.8465, -123.1452	XXO
H. icenoglei	GLR283	M	Puerto Peñasco, Estero Morúa, E of Playa Encanto, SONORA, MEXICO	31.273, -113.361	XXO
pugillis group					
H. pugillis	GLR236	M	Mt. Hopkins Road, Santa Cruz Co., ARIZONA, U.S.A.	31.689, -110.975	XXO
H. sp. (SUNGL)	GLR218	F	Amateur Astronomy Vista, Mt. Hopkins Rd, Santa Cruz Co., ARIZONA, U.S.A.	31.6759, -110.9289	uncertain
banksi group					
H. zapotecanus	GLR339	M	La Viuda restaurant, Chamela, JALISCO, MEXICO	19.5316, -105.0707	XXY
H. sp. (CHMLA)	GLR352	M	Estación de Biología Chamela, 400-650 m on Calandria Trail, JALISCO, MEXICO	19.5038, -105.0334	uncertain

Species	Voucher #	Sex	Locality	Latitude , Longitude	Species karyotype (male sex chromosomes only)
outgroups					
Pellenes cf. *levii*	GLR106	M	Mt. Baldy Road, BRITISH COLUMBIA, CANADA	49.1135, -119.2103	XXO
Evarcha prozjinski	GLR135	M	Silvermere Lake, Mission, BRITISH COLUMBIA, CANADA	49.166, -122.409	XXO
agilis group					
H. conjunctus	GLR234	M	Madera Canyon, near Proctor Road, Pima Co., ARIZONA, U.S.A.	31.7417, -110.8847	XXO
altanus group					
H. altanus	GLR180	M	East of Smoky Lake, ALBERTA, CANADA	54.112, -112.198	XXY
sp. (ROBRT) group					
H. sp. (ROBRT)	GLR346	M	Estación de Biología Chamela, Chachalaca Trail, JALISCO, MEXICO	19.496, -105.042	XXO
decorus group					
H. decorus	GLR132	M	Silvermere Lake, Mission, BRITISH COLUMBIA, CANADA	49.166, -122.409	XXXY
dorotheae group					
H. geronimoi	GLR267	M	Miller Canyon,Huachuca Mountains., Cochise Co., ARIZONA, U.S.A.	31.416, -110.276	XXY
fallax group					
H. cambridgei	GLR351	M	Bocanegra beach, Puerto Vallarta, JALISCO, MEXICO	20.670, -105.274	XXO
tranquillus group					
H. hirsutus	GLR080	M	Mt. Kobau Road, BRITISH COLUMBIA, CANADA	49.095, -119.610	XXO

Species	Voucher #	Sex	Locality	Latitude , Longitude	Species karyotype (male sex chromosomes only)
hallani group					
H. hallani	GLR209	M	near Arivaca, Pima Co., ARIZONA, U.S.A.	31.668, -111.245	XXY
paratus group					
H. paratus	GLR363	M	Starfish Beach, Punta Galeta, Isla Colon, BOCAS DEL TORO, PANAMA	9.40376, -79.8635	XXO

Appendix 2 Summary of RNA-sequencing and transcriptome assemblies for all species (except DNA-sequencing for *H.paratus*): total number of reads sequenced and reads remaining after trimming and filtering, % of reads mapped to mitochondrial and nuclear references, number of transcripts, and average coverage for all de novo and reference-based transcriptome assemblies.

Species	# of raw reads	# of reads after trimming	% of reads left after trimming	% of reads assembled to nuclear	% of reads assembled to mtDNA	%GC	# contigs 200-500bp	# contigs 500-1000bp	# contigs > 1000bp	Average nuclear coverage	Average mtDNA coverage	Total contigs
coecatus group												
H. festus	156,826,220	123,414,758	78.7	67.0	15.9	33	20,398	4,376	4,072	174	n/a	28,846
H. pyrrithrix	30,979,896	23,954,560	77.3	65.9	15.4	36	6,037	2,899	1,892	93	23,939	10,828
H. virgulatus	13,823,012	10,967,254	79.3	60.1	18.8	33	4,349	2,155	1,167	60	13,256	7,671
H. borealis	31,803,862	29,404,408	92.5	64.0	16.6	38	8,494	3,304	2,194	89	31,830	13,992
H. mexicanus	11,390,510	10,318,886	90.6	59.5	21.1	36	5,179	1,995	1,018	50	14,828	8,192
H. sp. (BLNDI)	11,052,058	9,278,850	84.0	59.0	19.6	35	8,439	2,132	1,095	57	11,904	11,666
H. captiosus	11,515,080	10,026,878	87.1	62.8	16.8	35	5,679	2,519	1,503	46	11,492	9,701
americanus group												
H. ophrys	169,057,044	123,988,153	77.6	73.1	15.9	33	37,287	8,804	5,052	111	n/a	51,143
H. americanus	10,974,358	10,107,232	92.1	70.3	11.9	37	5,656	2,348	1,000	53	7,721	9,004
H. tarsalis	10,563,210	9,497,728	89.9	63.6	11.5	37	4,562	2,117	992	52	7,095	7,671
H. sansoni	10,963,598	9,383,970	85.6	65.2	15.1	33	5,412	2,392	1,207	44	9,010	9,011
H. sp. (ESTU)	22,859,706	18,199,030	79.6	66.1	14.0	35	7,357	3,170	1,933	63	16,711	12,460

Species	# of raw reads	# of reads after trimming	% of reads left after trimming	% of reads assembled to nuclear	% of reads assembled to mtDNA	%GC	# contigs 200-500bp	# contigs 500-1000bp	# contigs > 1000bp	Average nuclear coverage	Average mtDNA coverage	Total contigs
clypeatus group												
H. clypeatus	19,732,586	18,838,386	95.5	64.8	15.4	37	5,984	2,881	1,689	78	19,106	10,554
H. aztecanus	6,974,386	6,223,514	89.2	92.5	0.2	35	3,464	1,526	777	49	7,293	5,767
H. sp. (SLCTY)	15,954,538	15,953,730	100.0	70.0	8.7	37	9,935	3,378	2,251	60	13,666	15,564
amicus group												
H. signatus	150,582,550	126,114,806	83.8	64.8	10.0	37	11,329	3,955	3,052	87	n/a	18,336
H. ustulatus	137,650,334	120,525,796	87.6	69.5	7.0	37	11,039	4,086	3,211	82	n/a	18,336
viridipes group												
H. jucundus	14,017,246	12,015,998	85.7	65.1	16.0	37	5,841	2,227	1,108	50	12,504	9,176
H. calcaratus maddisoni	9,522,580	8,691,730	91.3	56.0	25.3	35	3,583	1,770	768	54	14,496	6,121
oregonensis group												
H. icenoglei	16,345,750	13,819,638	84.5	71.7	9.3	37	3,623	2,438	1,132	90	7,720	7,193
H. oregonensis	13,721,754	10,598,794	77.2	62.9	16.0	37	5,197	2,288	1,215	54	11,821	8,700

Species	# of raw reads	# of reads after trimming	% of reads left after trimming	% of reads assembled to nuclear	% of reads assembled to mtDNA	%GC	# contigs 200-500bp	# contigs 500-1000bp	# contigs > 1000bp	Average nuclear coverage	Average mtDNA coverage	Total contigs
pugillis group												
H. sp. (SUNGL)	23,875,266	21,128,070	88.5	71.2	8.9	38	6,185	2,607	1,698	89	12,169	10,490
H. pugillis	14,485,868	13,625,478	94.1	68.4	16.0	37	4,890	2,363	1,241	74	13,985	8,494
banksi group												
H. sp. (CHMLA)	10,253,210	8,896,604	86.8	70.8	13.8	36	4,887	2,183	1,220	48	8,382	8,290
H. zapotecanus	23,804,988	22,550,394	94.7	50.0	10.3	40	7,446	2,912	1,581	65	15,759	11,939
outgroups												
Pellenes cf. *levii*	22,692,970	18,291,042	80.6	55.7	15.8	36	5,464	2,403	1,559	65	18,840	9,426
Evarcha prozjinsk	24,090,418	21,543,358	89.4	38.5	21.5	37	3,316	1,503	597	60	30,009	5,416
agilis group												
H. conjunctus	29,102,382	25,453,304	87.5	63.7	15.3	35	6,045	2,629	1,680	94	25,324	10,354
altanus group												
H. altanus	8,321,456	7,546,466	90.7	61.0	13.7	36	5,225	1,050	1,035	45	6,632	7,310

Species	# of raw reads	# of reads after trimming	% of reads left after trimming	% of reads assembled to nuclear	% of reads assembled to mtDNA	%GC	# contigs 200-500bp	# contigs 500-1000bp	# contigs > 1000bp	Average nuclear coverage	Average mtDNA coverage	Total contigs
H.sp. (ROBRT) group												
H. sp. (ROBRT)	6,051,598	5,308,674	87.7	63.1	17.1	35	2,255	745	746	43	6,188	3,746
decorus group												
H. decorus	14,842,184	11,274,830	76.0	56.7	20.3	34	5,059	2,501	1,187	48	15,069	8,747
dorotheae group												
H. geronimoi	18,321,104	15,484,070	84.5	67.8	13.7	36	5,027	2,545	1,261	79	13,571	8,833
fallax group												
H. cambridgei	13,076,744	12,432,290	95.1	64.9	16.9	36	4,837	2,596	1,366	65	14,220	8,799
tranquillus group												
H. hirsutus	8,551,110	7,444,104	87.1	58.1	14.9	36	3,834	1,610	794	42	7,517	6,238
hallani group												
H. hallani	16,041,252	13,578,488	84.6	63.2	16.5	36	5,679	2,589	1,501	57	14,366	9,769

Species	# of raw reads	# of reads after trimming	% of reads left after trimming	% of reads assembled to nuclear	% of reads assembled to mtDNA	%GC	# contigs 200-500bp	# contigs 500-1000bp	# contigs > 1000bp	Average nuclear coverage	Average mtDNA coverage	Total contigs
paratus group												
H. paratus	24,000,000	23,926,440	99.7	31.6	0.0004	30	8,437	561	111	64	60	9,109

Appendix 3 Significant Concordance Factors (CFs > 0.05) and their credibility intervals for the americanus clade.

Bipartitions number codes: 1= *H. signatus*, 2= *H.* sp. (ESTU), 3= *H. ophrys*, 4= *H. tarsalis*, 5= *H. sansoni*, 6= *H. americanus*

Bipartition	Supported clade	Sample-wide CF	95% credibility	Genome-wide CF	95% credibility	In primary concordance tree?	SD
{1,2,3,4|5,6}	*H. sansoni* + *H. americanus*	0.906	0.887-0.923	0.9	0.869-0.928	Y	0
{1,2|3,4,5,6}	*H. ophrys* + *H. tarsalis* + *H. sansoni* + *H. americanus*	0.508	0.476-0.541	0.505	0.456-0.555	Y	0
{1,2,5,6|3,4}	*H. ophrys* + *H. tarsalis*	0.501	0.464-0.536	0.498	0.446-0.550	Y	0
{1,2,4|3,5,6}	*H. ophrys* + *H. sansoni*/ *H. americanus*	0.305	0.272-0.339	0.303	0.257-0.352	N	0
{1,5,6|2,3,4}	*H.* sp. (ESTU) + *H. ophrys*/*H. tarsalis*	0.19	0.165-0.216	0.19	0.152-0.230	N	0
{1,3,5,6|2,4}	*H.* sp. (ESTU) + H. *tarsalis*	0.136	0.112-0.162	0.136	0.102-0.174	N	0
{1,2,3|4,5,6}	*H. tarsalis* + H. *sansoni*/H. *americanus*	0.136	0.112-0.162	0.136	0.102-0.174	N	0
{1,3,4|2,5,6}	*H.* sp. (ESTU) + *H. sansoni*/*H. americanus*	0.087	0.068-0.108	0.087	0.060-0.118	N	0
{1,4|2,3,5,6}	*H.* sp. (ESTU) + *H. ophrys* + *H. sansoni* + *H. americanus*	0.055	0.038-0.074	0.056	0.034-0.082	N	0

Appendix 4 Significant Concordance Factors (CFs > 0.05) and their credibility intervals for the VCC clade.

Bipartitions number codes: 1=*H. ophrys*, 2=*H. calcaratus,* 3=*H. jucundus,* 4= *H.* sp. (ROBRT), 5= *H. aztecanus,* 6= *H. clypeatus,* 7= *H. dossenus*, 8= *H. mexicanus*, 9=*H.* sp. (BLNDI), 10= *H. pyrrithrix*, 11= *H. virgulatus*, 12= *H. festus*, 13= *H. captiosus*, 14= *H. borealis*

Bipartition	Supported clade	Sample-wide CF	95% credibility	Genome-wide CF	95% credibility	In primary concordance tree?	SD
{1,2,3,4,8,9,10,11,12,13,14|5,6,7}	*clypeatus* group	0.571	0.553-0.588	0.571	0.525-0.617	Y	0.001
{1,2,3,4,5,8,9,10,11,12,13,14|6,7}	*H. clypeatus* + *H.* sp. (SLCTY)	0.552	0.518-0.582	0.552	0.498-0.605	Y	0
{1,4,5,6,7,8,9,10,11,12,13,14|2,3}	*H. calcaratus* +*H. jucundus*	0.494	0.472-0.515	0.494	0.446-0.542	Y	0.002
{1,2,3,4,5,6,7,8,9,10,11,12|13,14}	*H. captiosus* + *H. borealis*	0.487	0.460-0.513	0.487	0.437-0.538	Y	0.003
{1,2,3,4,5,6,7|8,9,10,11,12,13,14}	*coecatus* group	0.389	0.360-0.408	0.389	0.339-0.437	Y	0.005
{1,2,3,4,5,6,7,8,11,12,13,14|9,10}	*H.* sp. (BLNDI) + *H. pyrrithrix*	0.385	0.360-0.408	0.385	0.337-0.434	Y	0.005
{1,2,3,4,5,6,7,8,9,10,11|12,13,14}	*H. festus* + *H. borealis* + *H. captiosus*	0.331	0.311-0.350	0.331	0.287-0.377	Y	0.004
{1,4|2,3,5,6,7,8,9,10,11,12,13,14}	VCC clade, excluding *H.*sp. (ROBRT)	0.267	0.246-0.292	0.267	0.223-0.314	Y	0.001
{1,2,3,4,5,6,7,8|9,10,11,12,13,14}	*coecatus* group except *H. mexicanus*	0.267	0.244-0.288	0.267	0.224-0.312	Y	0
{1,2,3,4|5,6,7,8,9,10,11,12,13,14}	*clypeatus* + *coecatus* group	0.23	0.213-0.248	0.23	0.191-0.271	Y	0.005
{1,2,3,4,5,6,7,8,12,13,14|9,10,11}	*H.* sp. (BLNDI*)* + *H.pyrrithrix* + *H.virgulatus*	0.117	0.101-0.137	0.117	0.086-0.153	Y	0
{1,2,3|4,5,6,7,8,9,10,11,12,13,14}	*H.* sp. (ROBRT) + *clypeatus*/*coecatus* group	0.196	0.174-0.222	0.196	0.155-0.240	N	0.011
{1,2,3,8,9,10,11,12,13,14|4,5,6,7}	*H.* sp. (ROBRT*)* + *clypeatus* group	0.188	0.170-0.205	0.188	0.151-0.227	N	0.006
{1,2,3,4,7,8,9,10,11,12,13,14|5,6}	*H. aztecanus* + *H.* sp. (SLCTY)	0.175	0.149-0.201	0.175	0.134-0.218	N	0.002

Bipartition	Supported clade	Sample-wide CF	95% credibility	Genome-wide CF	95% credibility	In primary concordance tree?	SD
{1,2,3,4,5,6,7,8,9,10,11,14|12,13}	*H. festus* + *H. captiosus*	0.153	0.130-0.172	0.153	0.116-0.191	N	0.001
{1,2,3,4,6,8,9,10,11,12,13,14|5,7}	*H. aztecanus* + *H. clypeatus*	0.115	0.091-0.141	0.115	0.081-0.155	N	0.001
{1,2,3,4,5,6,7,8,9,10|11,12,13,14}	*H. virgulatus* + *H. festus* + *H. captiosus* + *H. borealis*	0.113	0.093-0.130	0.113	0.082-0.147	N	0.001
{1,2,3,4,5,6,7,11,12,13,14|8,9,10}	*H. mexicanus* + *H* .sp. (BLNDI) + *H. pyrrithrix*	0.104	0.087-0.120	0.104	0.075-0.136	N	0
{1,2,3,4,5,6,7,8,10|9,11,12,13,14}	*H*. sp. (BLNDI) + *H. virgulatus* + *H. festus* + *H. borealis* + *H. captiosus*	0.101	0.087-0.116	0.101	0.073-0.132	N	0.005
{1,4,5,6,7|2,3,8,9,10,11,12,13,14}	*viridipes* + *coecatus groups*	0.103	0.087-0.122	0.103	0.073-0.137	N	0.004
{1,2,3,4,5,6,7,8,10,12,13,14|9,11}	*H*. sp. (BLNDI) + *H. virgulatus*	0.099	0.083-0.116	0.099	0.071-0.132	N	0
{1,2,3,4,5,6,7,8,9,10,13,14|11,12}	*H. festus* + *H. virgulatus*	0.093	0.077-0.110	0.093	0.065-0.125	N	0.001
{1,2,3,4,5,6,7,8,11|9,10,12,13,14}	*H*. sp. (BLNDI) + *H. pyrrithrix* + *H. festus* + *H. borealis* +*H. captiosus*	0.089	0.070-0.110	0.088	0.059-0.124	N	0.008
{1,2,3,4,5,6,7,8,10,11|9,12,13,14}	*H*. sp. (BLNDI) + *H. festus* + *H. borealis* + *H .captiosus*	0.086	0.066-0.104	0.086	0.055-0.119	N	0.009
{1,2,3,4,5,6,7,8,9,12,13,14|10,11}	*H. pyrrithrix* + *H. virgulatus*	0.084	0.064-0.104	0.084	0.054-0.117	N	0.007
{1,2,3,4,5,6,7,8,10,11,13,14|9,12}	*H*. sp. (BLNDI) +H. festu	0.083	0.066-0.099	0.083	0.056-0.113	N	0.001
{1,2,3,4,5,6,7,8,9,10,11,13|12,14}	*H. festus* + *H. borealis*	0.081	0.066-0.099	0.081	0.054-0.112	N	0.002
{1,2,3,4,5,6,7,9,11,12,13,14|8,10}	*H. mexicanus* + *H. pyrrithrix*	0.071	0.056-0.085	0.071	0.046-0.099	N	0.002

Bipartition	Supported clade	Sample-wide CF	95% credibility	Genome-wide CF	95% credibility	In primary concordance tree?	SD
{1,3,4|2,5,6,7,8,9,10,11,12,13,14}	*H. calcaratus* + *clypeatus* group and *coecatus* group	0.068	0.052-0.083	0.068	0.043-0.096	N	0.001
{1,2,3,4,5,6,7,8,13,14|9,10,11,12}	*H.* sp. (BLNDI) + *H. pyrrithrix* + *H. virgulatus* + *H. festus*	0.061	0.048-0.074	0.061	0.039-0.086	N	0.003
{1,3|2,4,5,6,7,8,9,10,11,12,13,14}	*H. calcaratus* + *H.* sp. (ROBRT) + *clypeatus* group + *coecatus* group	0.06	0.048-0.075	0.06	0.038-0.087	N	0.005
{1,2,4|3,5,6,7,8,9,10,11,12,13,14}	*H. jucundus* + *clypeatus* group + *coecatus* group	0.059	0.044-0.074	0.059	0.037-0.086	N	0
{1,2|3,4,5,6,7,8,9,10,11,12,13,14}	*H. jucundus* + *H.* sp. (ROBRT) + *clypeatus* group + *coecatus* group	0.058	0.044-0.074	0.058	0.035-0.086	N	0.006
{1,2,3,6,7,8,9,10,11,12,13,14|4,5}	*H.* sp. (ROBRT) + *H. aztecanus*	0.056	0.043-0.070	0.056	0.035-0.082	N	0
{1,2,3,4,5,6,7,8,11,13,14|9,10,12}	*H.* sp. (BLNDI) + *H. pyrrithrix* + *H. festus*	0.054	0.041-0.068	0.054	0.032-0.080	N	0
{1,2,3,4,8|5,6,7,9,10,11,12,13,14}	*clypeatus* group and *coecatus* group, excluding *H. mexicanus*	0.054	0.041-0.066	0.054	0.032-0.079	N	0.002
{1,2,3,4,5,6,7,9,10,12,13,14|8,11}	*H. mexicanus* +*H. virgulatus*	0.052	0.039-0.066	0.052	0.031-0.078	N	0.002
{1,5,6,7,8,9,10,11,12,13,14|2,3,4}	*viridipes* group + *H.* sp. (ROBRT)	0.052	0.039-0.068	0.052	0.031-0.078	N	0.001

Appendix 5 Counts of shared alleles between *Habronattus* species used for D_{FOIL} tests

	Species 1	Species 2	Species 3	Species 4	Outgro-up	AAAAA	BBBBA	BBAAA	AABBA	ABABA	ABBAA	ABBBA	BAABA	BABAA	BABBA	BBABA	BBBAA	BAAAA	ABAAA	AAABA	AABAA
VCC clade																					
(A)	*aztecanus*	*clypeatus*	*mexicanus*	*pyrrithrix*	*ophrys*	1261335	16612	1950	1550	105	91	300	120	102	279	569	358	3017	2014	2951	4534
(B)	sp. (BLNDI)	*pyrrithrix*	*clypeateus*	*aztecanus*	*ophrys*	1305813	17312	2452	2108	55	55	248	37	52	188	300	346	2368	1846	3122	2023
(C)	sp. (SLCTY)	*clypeatus*	sp. (BLNDI)	*pyrrithrix*	*ophrys*	1505981	17856	2576	2813	42	33	160	21	30	124	231	176	1192	1235	1711	2373
(D)	sp. (SLCTY)	*clypeatus*	*mexicanus*	*pyrrithrix*	*ophrys*	*1399985*	16507	2264	1709	68	38	145	34	40	99	471	333	1072	1140	2806	4400
(E)	sp. (SLCTY)	*clypeatus*	*borealis*	*pyrrithrix*	*ophrys*	1598152	18920	2600	2072	80	39	144	31	65	127	439	350	1208	1324	3004	3711
(F)	sp. (SLCTY)	*clypeatus*	*borealis*	*pyrrithrix*	*ophrys*	*1344100*	17806	2068	1787	89	78	340	75	88	312	490	338	3218	2121	2860	3650
americanus group																					
(G)	*sansoni*	*american-us*	*ophrys*	*tarsalis*	*signatus*	1144994	33651	1257	678	34	49	74	21	37	63	266	468	375	562	1700	1325
decorus group with *americanus* group																					
(H)	*ophrys*	sp. (ESTU)	*zapotecan-us*	*decorus*	*signatus*	1433854	31209	6391	2939	189	225	347	231	167	431	1853	1313	3424	6131	9970	14493
(I)	*ophrys*	sp. (ESTU)	*cambridg-ei*	*decorus*	*signatus*	1401567	31044	6024	2502	176	211	361	210	157	444	1491	1561	3395	6013	10165	10913
(J)	*ophrys*	sp. (ESTU)	*oregonen-sis*	*decorus*	*signatus*	1389314	31290	5940	2377	172	184	360	177	176	478	1303	1703	3370	6048	10296	9094
(K)	*ophrys*	sp. (ESTU)	*jucundus*	*decorus*	*signatus*	1389556	31429	5766	2391	171	211	355	177	191	483	1232	1784	3346	5971	10300	9529
(L)	*ophrys*	sp. (ESTU)	*festus*	*decorus*	*signatus*	1475505	33310	6145	2575	191	252	392	194	229	513	1523	1973	3595	6462	11029	12866

Appendix 6 Counts of alleles shared between *Habronattus* species used for Patterson's D statistic tests.

	Species 1	Species 2	Species 3	Outgroup	ABBA	BABA	AAAA	AABA	ABAA	BAAA	BBAA	BBBA
americanus group (*H*.sp.(ESTU))												
(A)	*ophrys*	*tarsalis*	sp. (ESTU)	*signatus*	604	437	1460667	6699	2314	2323	2062	43518
(B)	*sansoni*	*americanus*	sp. (ESTU)	*signatus*	106	98	1314158	5972	763	465	2939	37177
decorus group with *americanus* group												
(C)	*zapotecanus*	*decorus*	*ophrys*	*signatus*	2276	1620	1498519	10627	10931	16117	3547	34071
(D)	*cambridgei*	*decorus*	*ophrys*	*signatus*	1868	1883	1466423	10183	11122	12084	3076	33903
(E)	*oregonensis*	*decorus*	*ophrys*	*signatus*	1637	2058	1454931	10087	11288	10140	2970	34321
(F)	*jucundus*	*decorus*	*ophrys*	*signatus*	1537	2138	1449870	9819	11189	10558	2961	34214
(G)	*festus*	*decorus*	*ophrys*	*signatus*	1906	2401	1551936	10659	12187	14479	3233	36798
H. sp. (ROBRT) and VCC clade												
(H)	sp. (SLCTY)	*jucundus*	sp. (ROBRT)	*ophrys*	531	1255	1128821	4047	4228	3236	1025	12437
(I)	*festus*	*jucundus*	sp. (ROBRT)	*ophrys*	1031	1473	1315652	6191	5550	7773	1985	16703
(J)	*festus*	sp. (SLCTY)	sp. (ROBRT)	*ophrys*	1051	476	1136547	4356	2939	5551	1541	13284

Appendix 7 Summary of D_{FOIL} results. p-value threshold for significance adjusted using a Bonferroni correction for multiple comparisons $p <$ 0.00074 (95% significance for 68 tests)

	Species 1	Species 2	Species 3	Species 4	Outgroup	total # sites	DFO	DFO *p*	DIL	DIL *p*	DFI	DFI *p*	DOL	DOL *p*	introgression
clypeatus and *coecatus* group															
(A)	sp. (BLNDI)	*pyrrithrix*	*clypeateus*	*aztecanus*	*ophrys*	1411233	0.061	6.E-02	0.027	4.E-01	-0.070	0.068	-0.116	2.E-03	none
(B)	sp. (SLCTY)	*clypeatus*	sp. (BLNDI)	*pyrrithrix*	*ophrys*	1564822	-0.086	4.E-02	-0.122	4.E-03	-0.058	0.227	-0.105	3.E-02	none
(C)	*aztecanus*	*clypeatus*	*mexicanus*	*pyrrithrix*	*ophrys*	1363153	-0.146	4.E-08	-0.155	7.E-09	-0.015	0.624	-0.026	4.E-01	*aztecanus*/*clypeatus* <-> *pyrrithrix*
(D)	sp. (SLCTY)	*clypeatus*	*mexicanus*	*pyrrithrix*	*ophrys*	1454728	-0.096	3.E-03	-0.167	2.E-07	-0.011	0.814	-0.166	4.E-04	clypeatus <-> pyrrithrix
(E)	sp. (SLCTY)	*clypeatus*	*borealis*	*pyrrithrix*	*ophrys*	1660568	-0.008	8.E-01	-0.141	6.E-06	0.113	0.011	-0.157	4.E-04	clypeatus <-> pyrrithrix
(F)	*aztecanus*	*clypeatus*	*borealis*	*pyrrithrix*	*ophrys*	1453126	-0.094	1.E-03	-0.132	4.E-06	-0.011	0.733	-0.055	8.E-02	none
americanus group															
(G)	*sansoni*	*americanus*	*ophrys*	*tarsalis*	*signatus*	1221642	0.267	6.E-16	0.235	1.E-12	0.019	0.737	-0.075	2.E-01	*sansoni*/*americanus* -> *ophrys*
decorus and *americanus* group															
(H)	*ophrys*	sp. (ESTU)	*zapotecanus*	*decorus*	*signatus*	*1559881*	*-0.161*	*0.E+00*	*-0.111*	*3.E-12*	*-0.010*	*0.688*	*0.116*	*4.E-06*	*ophrys -> decorus*
(I)	*ophrys*	sp. (ESTU)	*cambridgei*	*decorus*	*signatus*	1522609	-0.005	8.E-01	0.042	1.E-02	-0.003	0.899	0.110	1.E-05	none
(J)	*ophrys*	sp. (ESTU)	*oregonensis*	*decorus*	*signatus*	1511334	0.104	2.E-10	0.111	1.E-11	0.068	0.008	0.085	9.E-04	*ophrys*/sp. (ESTU) <-> *oregonensis*
(K)	*ophrys*	sp. (ESTU)	*jucundus*	*decorus*	*signatus*	1513001	0.140	0.E+00	0.153	0.E+00	0.064	0.010	0.097	1.E-04	none
(L)	*ophrys*	sp. (ESTU)	*festus*	*decorus*	*signatus*	1608166	0.097	1.E-10	0.109	6.E-13	0.054	0.024	0.083	5.E-04	none

Appendix 8 Summary of Patterson's D statistic results. p-value threshold for significance adjusted using a Bonferroni correction for multiple comparisons $p < 0.00074$ (95% significance for 68 tests)

Group	Species 1	Species 2	Species 3	Outgroup	total # sites	*D*	*p*	introgression
americanus group (*H*. sp. (ESTU))								
(A)	*ophrys*	*tarsalis*	sp. (ESTU)	*signatus*	*1526228*	*0.146*	*2.E-06*	*tarsalis* <-> sp. (ESTU)
(B)	*sansoni*	*americanus*	sp. (ESTU)	*signatus*	*1357252*	*0.115*	*8.E-02*	*none*
decorus group with *americanus* group								
(C)	*zapotecanus*	*decorus*	*ophrys*	*signatus*	1527650	0.174754707	0.E+00	decorus <-> ophrys
(D)	*cambridgei*	*decorus*	*ophrys*	*signatus*	1490550	0.003888889	8.E-01	none
(E)	*oregonensis*	*decorus*	*ophrys*	*signatus*	1478768	-0.113111986	2.E-11	oregonensis <-> ophrys
(F)	*jucundus*	*decorus*	*ophrys*	*signatus*	*1547434*	-0.149	0.E+00	*jucundus* <-> *ophrys*
(G)	*festus*	*decorus*	*ophrys*	*signatus*	1660536	-0.103	3.E-12	*festus* <-> *ophrys*
H. sp. (ROBRT) and VCC clade								
(H)	sp. (SLCTY)	*jucundus*	sp. (ROBRT)	*ophrys*	*1143143*	-0.405	0.E+00	sp. (SLCTY) <-> sp. (ROBRT)
(I)	*festus*	*jucundus*	sp. (ROBRT)	*ophrys*	*1415913*	-0.166	0.E+00	*festus* <-> sp. (ROBRT)
(J)	*festus*	sp. (SLCTY)	sp. (ROBRT)	*ophrys*	*1182634*	0.360	0.E+00	sp. (SLCTY) <-> *sp. (ROBRT)*