Systematic Entomology

Royal Entomological Society

Systematic Entomology (2022), **47**, 94–112 DOI: 10.1111/syen.12521

Describing biodiversity in the genomics era: A new species of Nearctic Cynipidae gall wasp and its genome

PEDRO F. P. BRANDÃO-DIAS[1], YUANMENG MILES ZHANG[2], STACY PIRRO[3], CAMILA C. VINSON[1], KELLY L. WEINERSMITH[1], ANNA K. G. WARD[4], ANDREW A. FORBES[4] and SCOTT P. EGAN[1]

[1]Department of BioSciences, Rice University, Houston, Texas, USA, [2]Systematic Entomology Laboratory, USDA-ARS, c/o National Museum of Natural History, Washington, District of Columbia, USA, [3]Iridian Genomes, Inc., Bethesda, Maryland, USA and [4]Department of Biology, University of Iowa, Iowa City, Iowa, USA

Abstract. Gall wasps (Hymenoptera: Cynipidae) specializing on live oaks in the genus *Quercus* (subsection *Virentes*) are a relatively diverse and well-studied community with 14 species described to date, albeit with incomplete information on their biology, life history and genetic structure. Incorporating an integrative taxonomic approach, we combine morphology, phenology, behaviour, genetics and genomics to describe a new species, *Neuroterus valhalla* **sp. nov.**. The alternating generations of this species induce galls on the catkins and stem nodes of *Quercus virginiana* and *Quercus geminata* in the southern United States. We describe both generations in the species' life cycle, and primarily use samples from a population in the centre of Houston, Texas, thus serving as an example of the undescribed biodiversity still present in well-travelled urban centres. In parallel, we present a draft assembly of the *N. valhalla* genome providing a direct link between the type specimen and reference genome. The genome of *N. valhalla* is the smallest reported to date within the tribe Cynipini, providing an important comparative contrast to the otherwise large genome size of cynipids. While relatively small, the genome was found to be composed of >64% repetitive elements, including 43% unclassified repeats and 11% retrotransposons. A preliminary ab initio and homology-based annotation revealed 32,005 genes, and a subsequent orthogroup analysis grouped 18,044 of these to 8186 orthogroups, with some evidence for high levels of gene duplications within Cynipidae. A mitochondrial barcode phylogeny linked each generation of the new species and a phylogenomic ultraconserved element (UCEs) phylogeny indicates that the new species groups with other Nearctic *Neuroterus*. However, both phylogenies present the genus *Neuroterus* in North America as polyphyletic.

Introduction

New species are described at an increasing pace (Costello et al., 2012, 2013), yet global diversity projections suggest we are far from a complete catalogue of the species on Earth (Scheffers et al., 2012; Stork, 2018). Notably, the rate of description of new species has increased even as the pace of species extinctions continues to rise (Dirzo et al., 2014). Therefore, much biodiversity is being lost prior to being described (Tedesco et al., 2014). Additionally, while undescribed species will occur more often in underexplored tropical biodiversity hotspots and other remote areas (Giam et al., 2012), they are certainly not restricted to such regions. The recent description of large land animals (e.g. Cozzuol et al., 2013), marine mammals (e.g. Yamada et al., 2019), and hundreds of insect species at a time (Srivathsan et al., 2019) testify to our ignorance of these underexplored ecosystems. However, there are new species still hiding in our proverbial backyards (e.g. Duran et al., 2019,

Correspondence: Scott P. Egan, Department of BioSciences, Rice University, Houston, TX 77005, USA. E-mail: scott.p.egan@rice.edu

Egan et al., 2017, Hartop et al., 2015, Jatnika et al., 2019, Samacá-Sáenz et al., 2020).

A major source of undescribed animal diversity includes galling arthropod systems and their associates, which make up complex food webs of minute insects dominated by hymenopterans (Forbes et al., 2016; Maia et al., 2018; Weinersmith et al., 2020). Total species richness in gall systems is estimated to be from 13 000 (Buhr, 1965) to over 210,000 species worldwide (Espirito-Santo & Fernandes, 2007), and this does not begin to address the estimated number of parasitoids attacking them (Forbes et al., 2018). Thus, many species remain to be formally described and Hymenoptera is now recognized as one of the most species-rich group of organisms on Earth (Forbes et al., 2018). More specifically, while Palearctic gall-former diversity has been more thoroughly described, much remains to be understood in other faunal regions (Penzes et al., 2018), including the Nearctic (Egan et al., 2018).

With over 1400 described species worldwide, Cynipidae is the second largest group of gall inducing insects (Ronquist et al., 2015). Over 1000 species belong to the tribe Cynipini, where the vast majority make galls on oaks (*Quercus*) and other related plants within Fagaceae. Cynipini have an unusual life cycle called heterogony or cyclical parthenogenesis, characterized by the alternation between a sexual and an asexual generation that develop in two morphologically and temporally dissimilar galls, commonly in different host tissues (Figure 1; Folliot, 1964; Pujade-Villar et al., 2001). This unusual life cycle implies that each cynipid gall-forming species actually produces two ecologically and functionally independent organisms, creating a unique taxonomic challenge. For instance, many species are currently only known from one of the presumed alternating generations, a common pattern within cynipid taxonomy even in the Palearctic (Stone et al., 2008), while many others have had their alternating generations – which can often look quite different – described as two separate species (e.g., *Belonocnema* asexual and sexual generations from Florida, reviewed by Zhang et al., 2021).

Gall wasps on live oaks (*Quercus*; subsection *Virentes*) make up a well-studied community whose hosts are distributed in terrestrial habitats surrounding the Gulf of Mexico, including the southeastern United States, eastern Mexico and Central America, western Cuba and a geographically disjunct live oak species in southern Baja, Mexico (Cavender-Bares et al., 2015; Manos & Hipp, 2021; Muller, 1961). This group includes the iconic southern live oak (*Quercus virginiana* Miller), which is widely used in urban landscaping and is an economically important nursery crop (USDA, 2021). These oaks harbour an abundant and relatively diverse community of galling insects and associated natural enemies (Egan et al., 2013, 2018), including over a dozen galling cynipid species (Table 1). Each of these gall-formers support a complex interconnected food web, complete with inquilines, parasitoids, hyperparasitoids and commensal associates (e.g., Forbes et al., 2016; Weinersmith et al., 2020), potentially totalling over 100 species found on a single tree. Although many species in this system were described as early as in the 19th century, many of the live oak gall-former species are only known from one of the two generations

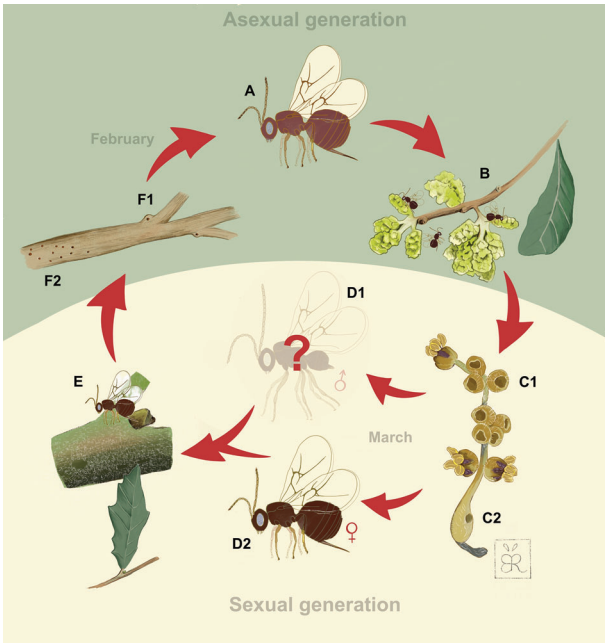

Fig. 1. *Neuroterus valhalla* life cycle. (A) Stem node (asexual) generation female; (B) oviposition in developing catkin buds; (C) N. valhalla's oval gall on the catkin inflorescence (C1), which are not to be mistaken with *Andricus quercuslanigera*'s fusiform galls on the stalk of the catkin (C2); (D) unknown sexual generation male (D1), catkin (sexual) generation female (D2); (E) oviposition in stem nodes; (F) N. valhalla's cryptic galls on stem nodes (F1), not to be mistaken with *Bassettia pallida*'s internode clustered cryptic galls (F2). Green background highlights the asexual (stem node) generation, while yellow background highlights the sexual (catkin) generation. Illustration by Barbara Rossi.

(Table 1). These taxonomic challenges can cascade up trophic levels to some of the gallers' natural enemies, with species level identification of many taxa being highly challenging due to a combination of historical species descriptions lacking host association information (i.e. species described solely on museum specimens with no natural history information), and high levels of cryptic morphological diversity in some groups (e.g., *Eurytoma*; Zhang et al., 2014; *Synergus*; Ward et al., 2020).

One solution to these various problems is integrative taxonomy, the incorporation of multiple lines of evidence in support of new species hypotheses, including natural history, ecology, biogeography, taxonomy, behaviour, genetic and genomic data (e.g. Dayrat, 2005; Deans et al., 2012; Duran et al., 2019, 2020; Godfray, 2002; Padial et al., 2010). For example, while species descriptions remain rooted in morphology, there is a strong incentive to include a DNA barcoding sequence with the species descriptions (Cook et al., 2010; Hebert & Gregory, 2005; Padial & De la Riva, 2007). Accordingly, DNA barcoding has become an integral part of a thorough species description (e.g. Cozzuol et al., 2013; Egan et al., 2017). This popularization of DNA barcoding has facilitated species identification by nonspecialists, something increasingly relevant in the context of declining taxonomic expertise (Ebach et al., 2011; Orr et al., 2021; Pearson et al., 2011), and has proved useful as a complement

Table 1. Cynipid gall wasps reported from the same live oak hosts as *Neuroterus valhalla*

Species	Hosts	Both Gen	References
Andricus quercusfoliatus	Qv, Qg	No	Abrahamson et al., 1998; Ashmead, 1881
Andricus quercuslanigera	Qv, Qg	Yes	Abrahamson et al., 1998; Ashmead, 1881; Hood et al., 2018
Bassettia pallida	Qv, Qg	No	Abrahamson et al., 1998; Ashmead, 1896
Belonocnema fossoria	Qg	Yes	Weld, 1921; Zhang et al., 2021
Belonocnema kinseyi	Qv	Yes	Weld, 1921; Zhang et al., 2021
Belonocnema treatae	Qv	Yes	Mayr, 1881; Zhang et al., 2021
Callirhytis quercusbatatoides	Qv, Qg	No	Abrahamson et al., 1998; Ashmead, 1881
Disholcaspis cinerosa	Qv	Yes	Bassett, 1881; Frankie et al., 1977
Disholcaspis quercusvirens	Qv, Qg	Yes	Ashmead, 1881; Melika et al., 2013
Loxaulus pattersoni	Qv	No	Melika & Abrahamson, 2000
Neuroterus christi	Qv, Qg	No	Melika & Abrahamson, 1997
Neuroterus quercusminutissimus	Qg	No	Ashmead, 1885, Abrahamson et al., 1998
Neuroterus saltatorius var. texanus	Qv	Yes	Kinsey, 1923
Neuroterus valhalla	**Qv, Qg**	**Yes**	***In this study***
Odontocynips nebulosa	Qv	No	Kieffer, 1910; Wilson et al., 2000

Note: Host associations reported do not include live oak host species outside of Qg and Qv.
Abbreviation: Qg, *Quercus geminata*; Qv, *Quercus virginiana*.
Neuroterus valhalla is highlighted in bold.

to morphology in species discrimination, particularly in groups with considerable cryptic morphological diversity (e.g. Smith et al., 2008; Xu et al., 2015). However, COI has been shown to be sometimes unreliable for species identification due to incomplete lineage sorting (Rokas et al., 2003) and hybridization (Nicholls et al., 2012), which may result in mitonuclear discordance and consequent misleading results from COI alone (Cook et al., 2002; Zhang et al., 2021).

Fortunately, advances in molecular biology are accelerating with numerous improvements in sequencing and computing technology (Buermans & Den Dunnen, 2014; Goodwin et al., 2016). The consequent lower costs of genome sequencing have taken us from the first vertebrate genome sequence, our own (International Human Genome Sequencing Consortium, 2001; Venter et al., 2001), to the possibility of sequencing the genome of every known vertebrate species (Rhie et al., 2021), thousands of arthropods (i5K Consortium, 2013), or millions of unicellular eukaryotes (Lewin et al., 2018). With this, biological research is fundamentally growing into the genomics era. Taxonomy has embraced this change, and genomic tools allowed taxonomists to discover new cryptic species, as well as untangle the evolutionary history of challenging groups (e.g. Duran et al., 2020; Hassemer et al., 2019). Further, as genome sequencing technologies become easier, less expensive and more widely available, we are approaching the point where, as resources allow, it might be as easy to sequence a genome as it is to describe a species' morphology. Thus, genome sequencing may become a fundamental piece of the integrative evidence for the description of new species and complement formal morphological description. For example, the incorporation of genome sequencing in concert with new species descriptions is a standard practice in microbiology (e.g. Bányai et al., 2017; Lorch et al., 2018).

Here, we describe a new species of Cynipidae, *Neuroterus valhalla* **sp. nov.**, which induces galls on live oaks on the *Virentes* group (*Q. virginiana* and *Quercus geminata* small). We include a description of its life cycle, with both alternating heterogonic generations, and a draft assembly and annotation of its genome, which we believe reflects the future of animal species descriptions, and the first new insect species description to include its genome in concert with the morphological description. To our knowledge, this is only the second new animal species description to do so, where the first was a newly discovered sibling species of *Caenorhabditis elegans* (Kanzaki et al., 2018). We then demonstrate the utility of genomic data by harvesting UCE loci from the *N. valhalla* genome to improve published phylogenomic relationships of Cynipini. Additionally, *N. valhalla* is described from a population found in the heart of Houston, Texas, USA, one of the most populated cities in North America. Despite the minute size of its galls, *N. valhalla* appears to number in the tens of millions every year when it emerges from galls as an adult, making it an example of overlooked and undescribed biological diversity found in the centre of a well-known urban centre.

Methods

Plant tissue sampling and insect rearing

We sampled catkins (inflorescence composed of male staminate flowers) of southern live oaks (*Q. virginiana*) throughout the flowering season between February and March of 2018, 2019 and 2020 at the Rice University campus (29.718 N, 95.400 W). Due to the minute size of galls, we sampled flowering structures indiscriminately using gardening shears. We checked samples for emerging insects three times a week until May, when emergence stopped. Then, from May 2020 to February 2021, we checked samples occasionally but observed no further emergences. The total number of collected flowers

was then extrapolated by weight given an average flower count and weight of 10 random sub-samples.

Then, to confirm the location of the alternative generation, in January and February of 2020 we focused on trees in which we had detected the species emerging from catkins in the previous year. Following the observation of emergence holes in stem nodes, we sampled branches where there were no observable alternative gall structures from other known species (e.g. the Cynipidae *Bassettia pallida* Ashmead, *Callirhytis quercusbatatoides* [Ashmead], and Cecydomiidae *Arnoldiola atra* Gagné), and trimmed away leaves and buds from them.

We kept all plant samples (catkins and branches) in a greenhouse and inside our gall rearing chambers: glass mason jars mounted with an upside-down funnel on the lid, leading into a *Drosophila* vial where emerging insects get trapped until collected. Samples were checked three times a week for emerging insects until May, when emergence slowed down, and occasionally after that. We then manually sifted samples with gall wasp emergence under stereoscopic microscopes to look for gall structures.

mtDNA barcoding

For DNA barcoding of individual insects, we extracted DNA from whole specimens using the DNeasy Blood and Tissue Kit (Qiagen, Germany) in accordance with the manufacturer's protocol with the addition of a pestle crushing step prior to incubation. The mitochondrial cytochrome oxidase I (COI) gene was amplified using the KAPA Taq ReadyMix (Sigma Aldrich, St. Louis, Missouri) and the primers LEP-F: 5′-TAAACTTCTGG ATGTCCAAAAAATCA-3′ and LEP-R: 5′-ATTCAACCAATA CATAAAGATATTGG-3′ (Hebert et al., 2004). Reactions consisted of 0:45 s at 94°C for initial denaturation, followed by 35 cycles of 00:15 s at 94°C for denaturation, 00:15 s at 58°C for annealing and 00:30 s at 72°C for extension. Amplicons were Sanger sequenced at the University of Arizona Genetics Core. Results were aligned and manually curated using MEGA 10 (Kumar et al., 2018).

mtDNA COI phylogeny

To build the phylogeny with the mtDNA COI barcodes, we downloaded all available *Neuroterus* and representative Cynipini COI sequences from GenBank, with the addition of extracted COI barcode locus of the *Neuroterus distortus* Bassett UCE data from Blaimer et al. (2020) using PHYLUCE v1.6.8 (Faircloth, 2015). Sequences were aligned using MAFFT v7.480 (Katoh & Standley, 2013), and the maximum-likelihood phylogeny was inferred using IQ-TREE v2.03 (Minh et al., 2020) under the K3Pu + F + I + G4 as selected by ModelFinder (Kalyaanamoorthy et al., 2017). Nodal support was assessed using 1000 nonparametric bootstraps, with values ≥90 considered to be strongly supported. The output trees were visualized in R v.4.0 (R Core Team, 2020) using the packages ggtree v.2.2.0 and treeio v.1.12.0 (Wang et al., 2020; Yu et al., 2017). Accession numbers and details on the sequences used in the phylogeny are available in Table S6, and the sequence alignment is available in the dryad repository (https://doi.org/10.5061/dryad.zgmsbccc8).

Morphological study

For the formal morphological description, ethanol-preserved specimens were dehydrated through increasing concentrations of ethanol and transferred to hexamethyldisilazane (HMDS) (Heraty & Hawks, 1998) before point-mounting. The specimens were identified using Kinsey (1923), in comparison with more contemporary species descriptions of Nearctic *Neuroterus* species (Medianero & Nieves-Aldrey, 2017; Melika & Abrahamson, 1997; Pujade-Villar et al., 2014, 2016, 2017, 2018). We took scanning electron microscope (SEM) images with a Hitachi TM3000 (Tungsten source). Body parts of disarticulated specimens were adhered to a 12.7 × 3.2 mm Leica/Cambridge aluminium SEM stub by a carbon adhesive tab (Electron Microscopy Sciences, #77825–12). Stub-mounted specimens were sputter coated with gold-palladium using a Cressington Scientific 108 Auto from multiple angles to ensure complete coverage (∼20–30 nm coating). Coloured images were obtained with a Canon 60D DSLR, with a Canon MP-E 65 mm F/2.8 Macro photo lens or a Mitutoyo M Plan Apo 10× objective mounted on to the Canon EF Telephoto 70–200 mm zoom lens, and the Canon MT–24EX Macro Twin Lite Flash (Tokyo, Japan) with custom-made diffusers to minimize hot spots. Images saved as TIF, and focus stacked using Zerene Stacker v1.04. Image editing was done in Adobe Photoshop and plate layout in Inkscape.

We follow Buffington et al. (2020); Liljeblad & Ronquist (1998); Melika (2006) for terminology on Cynipidae morphological structures and abbreviations for fore wing venation, and Harris (1979) for patterns of cuticular sculpture. The following measurements and abbreviations were used: F1–Fn, the first and the following flagellomeres; POL (postocellar distance), the distance between the inner margins of posterior ocelli; OOL (ocellar-ocular distance), the distance from the outer margin of lateral ocellus to the inner margin of compound eye; LOL (lateral-ocular distance), the distance between lateral and frontal ocellus; transfacial line, distance between inner margins of compound eyes measured across the toruli; width of radial cell, measured as the distance between the upper margin of the fore wing and the Rs vein. Voucher specimens are deposited at NMNH, and the research collection of the Egan lab (Rice University, Houston, USA).

Tissue preference tests

We conducted a tissue preference test to verify in which tissue the females emerging from the catkin generation would oviposit. Freshly emerged insects (up to 24-h old adults) were placed into 90 mm petri dish arenas containing six freshly collected *Q. virginiana* tissue pieces, placed in randomized order within

the arena: newly developing leaf, fully mature leaf, new-growth offshoot (stem node with an axillary bud), old-growth offshoot (stem node where axillary bud has developed into a branch), bud tissue and catkins. Light was set perpendicular to petri dish. Insects were then observed for 30 min, during which we recorded their location and behaviour within the arena every 2 min. All insects emerging during a 2-week interval were used (n = 33), which were all females. After observations were finished, individuals were placed within 96% ethanol vials and their identity was morphologically confirmed under dissecting microscopes. One observation was excluded from final analysis because the insect died during the behavioural trial.

Genome sequencing, assembly and annotation

To sequence *N. valhalla*'s genome, we extracted whole genomic DNA from a single female from the catkin generation using the DNeasy Blood and Tissue Kit (Qiagen) as described above. A paired-end sequencing library was then constructed with the Illumina TruSeq kit (Illumina, San Diego, California) using the standard adapters, and sequencing was performed at Genewiz (New Jersey, USA) on an Illumina X-Ten sequencing platform. Raw data were submitted to GenBank Short Read Archive (Accession number SRX7007139). Then, fastq files were filtered and trimmed by Trimmomatic v.0.39 (Bolger et al., 2014), first removing sliding windows of five nucleotides with quality average below 20, followed by a hard-trailing trim of nucleotides with quality below 15. Trimmed reads were then assembled *de novo* by SPAdes v3.14.0 (Bankevich et al., 2012) using kmer-sizes of 27, 37, 47, 57, 67, 77, 87, 97, 107, 117 and 127. Assembly quality and completeness were accessed by quast v5.0.2 (Gurevich et al., 2013) and BUSCO v4.0.6 (Simão et al., 2015).

For annotation, repetitive sequences were inferred using RECON (Bao & Eddy, 2002), RepeatScout (Price et al., 2005) and LTR_retriver (Ou & Jiang, 2018), as applied by RepeatModeler (Smit et al., 2015). Subsequently these sequences were classified with RepeatClassifier according to RepBase v25.04 (Bao et al., 2015), and representation of these sequences in the genome were accessed by RepeatMasker v4.0 (Smit et al., 2015), which also masked the assembly for subsequent gene annotation. Augustus (Stanke et al., 2006) was applied for ab initio gene prediction, using *Nasonia vitripennis* (Walker) (Werren et al., 2010) training parameters, the closest organism with a training set available. For protein-based annotation, we used a database containing all known genes from the most closely related organism with an annotated genome, *Belonocnema kinseyi* Weld (GenBank 17 056 478), and used Exonerate v.2.2 (Slater & Birney, 2005). Both annotations were then combined using EvidenceModeler v1.1.1 (Haas et al., 2008), with equal weight to either approach. The final .gff file was then filtered and analysed with gFACs v1.0 (Caballero & Wegrzyn, 2019) for gene statistics. Annotation files for both repetitive sequences and genes are available in the dryad repository (https://doi.org/10.5061/dryad.zgmsbccc8).

Gene ortholog analysis was performed using Orthofinder (Emms & Kelly, 2019). The analysis encompassed nine annotated genomes, including *Belonocnema kinseyi*; every proteome available in UniProt database belonging to the Infraorder Parasitoida (three total, including above mentioned *Neaethus vitripennis*); *Apis mellifera* L. and *Drosophila melanogaster* Meigen for their status as most well characterized insect genomes; and *Arabidopsis thaliana* (L.) Heynh. and *Homo sapiens* L. as outgroups.

Ultraconserved elements phylogeny

To build the UCE phylogeny, we downloaded the UCE assemblies from all Cynipini and selected outgroups from other tribes from Blaimer *et al.* (2020) and assembled the contigs using SPAdes. Additionally, UCE loci were extracted from all available Cynipini genomes on NCBI not already present in Blaimer *et al.* (2020) dataset, including *N. valhalla*, following tutorial III of the PHYLUCE pipeline, using the Hym-V2P probe set developed by Branstetter *et al.* (2017). All assemblies were aligned using MAFFT and trimmed using Gblocks v0.91b-2 (Castresana, 2000) using the following settings: b1 = 0.5, b2 = 0.5, b3 = 12, b4 = 7. Then, we used Spruceup 0.95 lognormal distribution or manual cut off of select samples to remove any potentially misaligned regions as they can produce exaggerated branch lengths (Borowiec, 2019). We selected the 50% complete matrix as the final dataset and inferred the maximum-likelihood phylogeny using IQ-TREE using best models for each locus selected by ModelFinder. To assess nodal support, we performed 1000 ultrafast bootstrap replicates (UFBoot2, Hoang et al., 2017), along with '-bnni' to reduce risk of overestimating branch supports; and a Shimodaira-Hasegawa approximate likelihood-rate test (SH-aLRT, Guindon et al., 2010) with 1000 replicates. Only nodes with support values of UFBoot2 $\geq$ 95 and SH-aLRT $\geq$80 were considered robust. All UCE sequences obtained are available in the dryad repository (https://doi.org/10.5061/dryad.zgmsbccc8).

Results

Discovery of new species and linking generations

Neuroterus valhalla **sp. nov.** was initially discovered incidentally while sampling catkins of *Q. virginiana* at Rice University, in Houston, Texas. We were originally looking for the sexual generation of *Andricus quercuslanigera* (Ashmead), another cynipid gall-former of *Q. virginiana* (Hamel, 1973; Hood et al., 2018). When we DNA barcoded individual catkin emergents, we found that some of the insects had sequences very different from *A. quercuslanigera* and belonged to a unique mitochondrial lineage (Figure 2, note divergent sequences marked in blue and grey; Egan et al., 2018; Hood et al., 2018). Following this, we re-examined emerging insects, and discovered two distinct morphotypes of female cynipid wasps – one was *A. quercuslanigera*, as expected, and the other was an

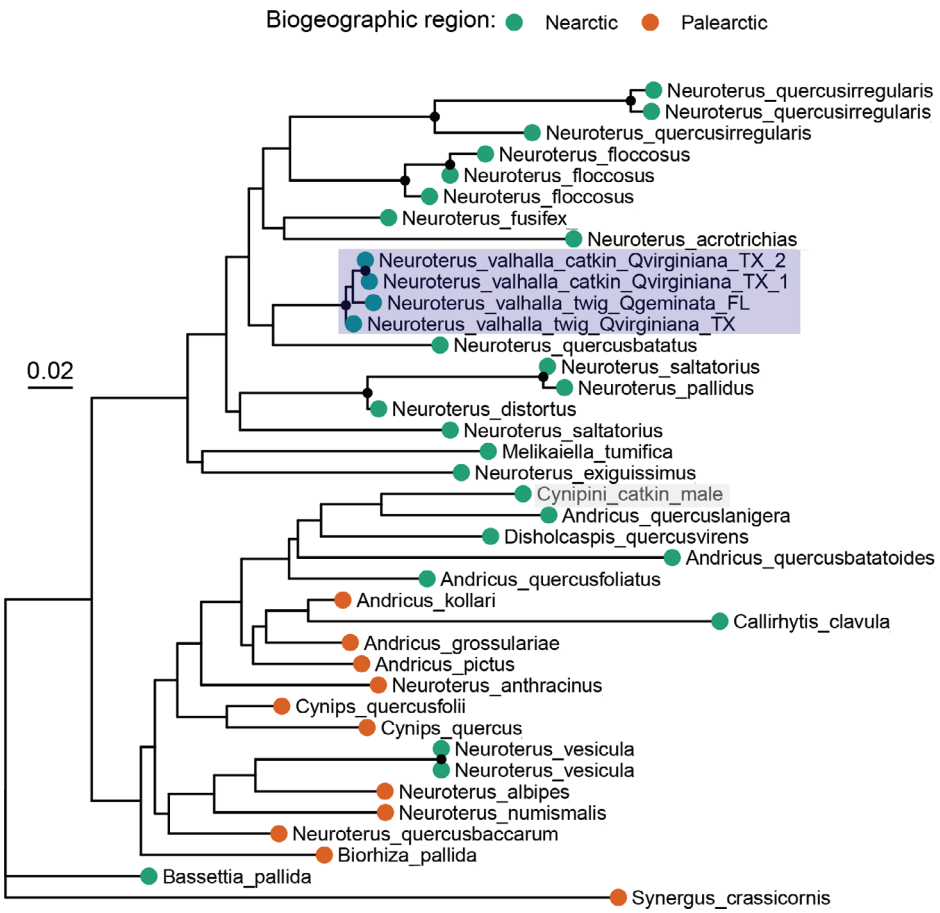

Fig. 2. Maximum-likelihood COI phylogeny of Cynipini. Strongly supported nodes (95% UFBoot values) are shown with black dots. Branch lengths represent genetic distance. Blue = *N. valhalla* sequences; Yellow = catkin-emerging male (*A. quercuslanigera*).

unknown female morphospecies. These novel females were then called *Neuroterus valhalla* (Figure 3a) since they were noticeably smaller with lighter legs than *A. quercuslanigera* and exhibited numerous unique features that separate the genus *Neuroterus* from *Andricus* in North America, including the absence of notauli and scutoscutellar suture (see morphlogical details below).

Although females could be separated into two groups, males emerging from catkins were all identical, and were identified as *A. quercuslanigera* by the presence of transscutal articulation, which is absent in *Neuroterus* and present in *Andricus*. Nonetheless, in the search for *N. valhalla* males, we DNA barcoded three males with particularly faint notauli and reduced size and confirmed them to be *A. quercuslanigera* (Figure 2, grey). Therefore, males of *N. valhalla* remain unknown, despite 117 female emergents among catkins samples across 3 years of sampling efforts.

In parallel, while sampling for another cynipid on *Q. geminata* in Florida, *Bassettia pallida* Ashmead, we found two very distinct cynipid wasps among several *B. pallida* (Weinersmith et al., 2020), which keyed to *Neuroterus*. *Bassettia pallida* galls are cryptic swellings of internode branches (Figure 1F2), and the presence of a second Cynipini wasp suggested we had inadvertently sampled a second cryptic gall on such branches (Figure 1F). We followed with DNA barcoding of these unknown cynipid wasps, expecting them to be one of the unknown sexual generations of the previously described sympatric species (Table 1). However, we found that the sequence to these wasps matched that of the catkin emergents from *Q. virginiana* in Texas (Figure 2, blue). Therefore, these insects likely represented *N. valhalla*'s alternating generation despite being initially identified in geographically distant populations (878 km apart) and on two distinct, but sister species of host oak trees.

To confirm the entire life cycle within the same host tree species, we visually inspected the *Q. virginiana* trees on Rice campus with the highest density of *N. valhalla* catkin emergence. We noticed that there were abundant emergence holes in stem nodes (Figures 1F1 and 3e), which after close inspection and dissection in the laboratory, were revealed to have the characteristic internal oval structure of a cynipid gall (see Weinersmith et al., 2020). We then sampled year-old stem nodes from trees, even though no gall structures were obviously visible, and 14 total cynipid wasps, which keyed to *Neuroterus* emerged from the collected material (Figure 3c) with a phenology

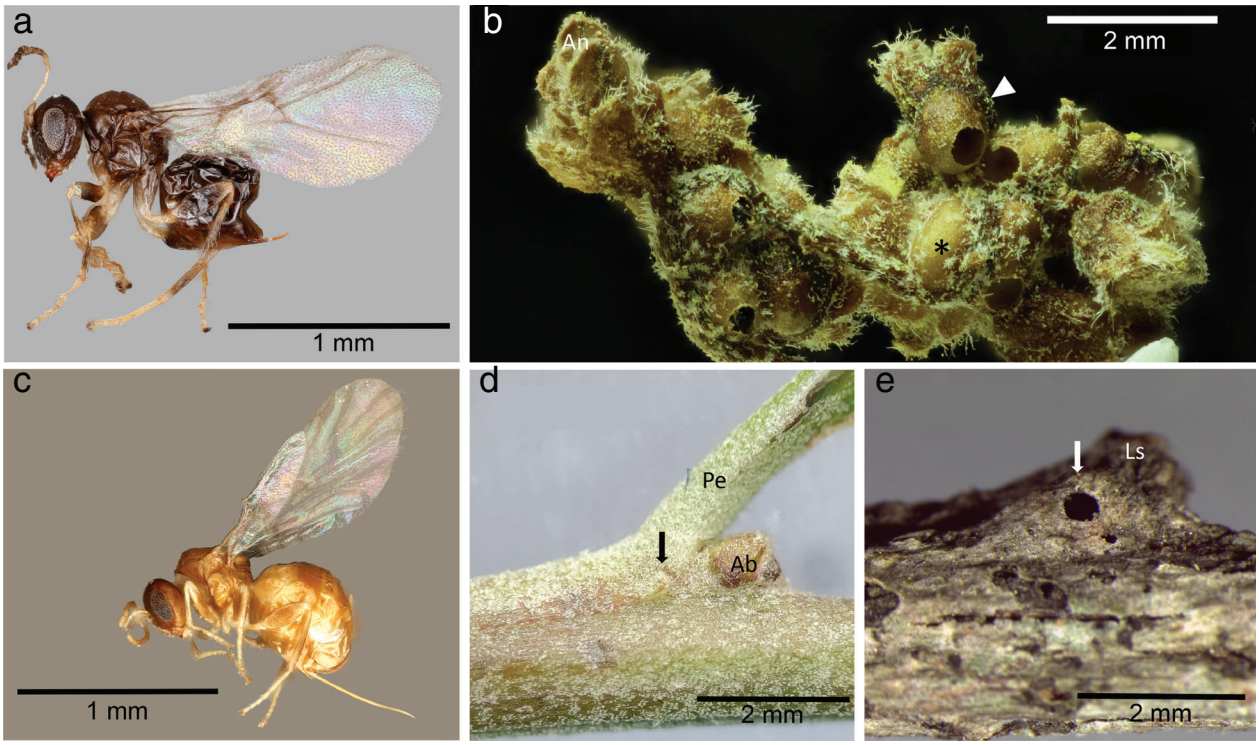

Fig. 3. Neuroterus valhalla. (a) Catkin (sexual) generation female holotype specimen habitus; (b) catkin gall cluster. White arrowhead = gall with emergence hole, black asterisk = gall without emergence hole, An = anther in healthy staminate flower; (c) Stem node (asexual) generation female habitus. Note that this specimen lost colour and in vivo looks darker; (d) stem node prior to gall development. Note the oviposition car (black arrow), which is a slight swell of the tissue with a black dot, Ab, axillary bud; Pe, petiole of adjacent leaf. (e) Stem node gall; Ls, leaf scar; White arrow, gall with emergence hole.

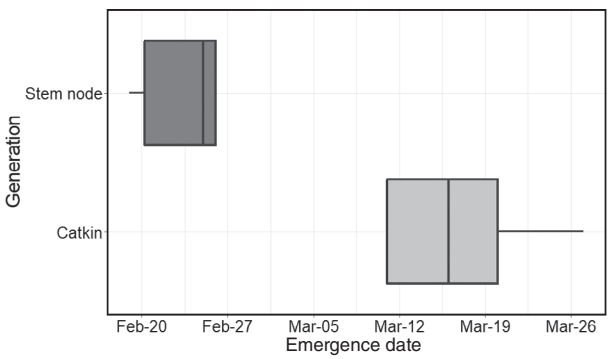

Fig. 4. Emergence Phenology of *Neuroterus valhalla*. Boxplot of the emergence date of each generation (n = 14 stem node generation, n = 117 catkin generation). Boxes indicate 25 and 75 percentiles, horizontal bars the range and vertical internal bar indicates the mean emergence date.

compatible with its oviposition into the simultaneously developing catkins (Figure 4). We barcoded one of these specimens, which confirmed that both the stem node emerging and catkin-emerging females were indeed conspecifics and grouped with the individual found earlier in *Q. geminata* (Figure 2, blue).

In the tissue preference tests of catkin-emerging females, we noted oviposition behaviour on three occasions: twice in year-old stems, once in new stems and never on offered buds, leaves or catkins (Figure S1). In the new stem gall, we observed an oviposition scar in the tissue following the behavioural observation (Figure 3d) in a location analogous to that observed in developed stem node galls (Figure 3e). These behavioural observations complemented the cross-generation link confirmed by DNA barcoding (Figure 2). Adult gallers were also often found among leaves and catkins, but they were mostly stationary, and no oviposition was ever observed (Figure S1).

Morphological description

Neuroterus valhalla **sp. nov**. Brandão-Dias, Zhang, Weinersmith, Forbes & Egan

Diagnosis: *Neuroterus valhalla* keys to couplet 6 in Kinsey (1923) in the subgenus *Diplobius*, which can be recognized by the presence of malar sulcus, 13 antennal segments, absence of parapsidal grooves, simple tarsal claws or with a very short tooth, and induces polythalamous leaf/stem galls and monothalamous anther galls. Only *N. valhalla* and three other *Neuroterus* species (*N. floricola* Kinsey, *N. verrucum* Pujade-Villar, and *N. fusifex* Pujade-Villar & Ferrer-Suay) are known to induce catkin or stem galls, all of which have alutaceous to delicate coriaceous mesoscutum (Kinsey, 1923; Pujade-Villar

et al., 2014, 2016). *Neuroterus valhalla* is the only species recorded from oak section *Virentes*, while the other three are found on oak section *Quercus*. *Neuroterus floricola* Kinsey, which are only known from the sexual generation and induces simple catkin galls similar to *N. valhalla* but on *Quercus douglasii* Hook. & Arn. in California, can be differentiated based on the presence of very faint anterior parallel lines in *N. floricola* which are absent in *N. valhalla* (Kinsey, 1923). The two Mexican species are *N. verrucum*, which is only known from the asexual generation that induces cryptic stem galls on *Quercus laeta* Liebm., and *N. fusifex*, which is only known from the sexual generation also found on *Q. laeta* in Mexico, induces multilocular, ovoid catkin galls. It is possible these two Mexican species are alternate generations of each other given the morphological and host use similarities, but a revision of the genus is beyond the scope of this paper. Nevertheless, *N. valhalla* differs from both Mexican species by the indistinct epistomal sulcus and clypeo-pleurostomal line and the transfacial distance slightly longer than the height of eye.

Material examined. Holotype: USA, TX, Houston, Rice University Campus, 29.718 N, 95.400 W, 11.III.2020 (Egan Lab Leg.), Em. 18.III.2020, Catkin gall on *Quercus virginiana*. USNMENT01448617, paratypes same locality as holotype 3♀, USNMENT01558416, 01558515, 01558518. Same locality and data as holotype, Em. 26.II.2020, gall in nodes of branches, 3♂, asexual generation, USNMENT01558623, 01558269, 01558597.

Sexual generation

Females

Body length = 1.1 mm (n = 4).

Colour: Body brown. Mandibles light yellow, antennae and apices of femur, basal area of tibia and whitish, tarsomeres dark brown (Figure 3a).

Head: Transverse in dorsal view, 3.0× as wide as high in front view and slightly wider than mesosoma (Figure 5a). Lower face alutaceous to smooth, with sparse setae, without striae radiating from clypeus. Gena 1.5× as wide as transverse diameter of eye, not visible in frontal view; malar space 0.3× as long as eye height, malar sulcus present; mandibles tridentate. Ocellar area slightly elevated; POL:OOL:LOL ratio 2.5:1:1.5. Transfacial distance 1.1× the height of eye; diameter of torulus (including rims) 1.1× that of intertorular distance; inner margins of eyes very slightly converge ventrally. Clypeus small, rounded, alutaceous in the centre, smooth lateral. Anterior tentorial pits distinct, epistomal and clypeo-pleurostomal line absent (Figure 5a). Frons, vertex and interocellar area alutaceous, shiny and glabrous. Occipital carina absent. Antenna (Figure 5d) with 13 antenomeres; pedicel rounded; F1 1.6× as long as F2, not enlarged or curved; ratio of antennal segments: 1.7:1.1:1.9:1.4:1.3:1.1:1.3:1.4:1.2:1.2:1.0:1.0:1.7; placodeal sensilla on F3-F13, and erect setae in all antennal segments.

Mesosoma: Around 1.1 times as long as high in lateral view, glabrous (Figure 5b). Pronotum alutaceous and shiny. Mesoscutum 1.1× as long as wide in dorsal view, weakly alutaceous, smooth in the centre, with very few sparse setae laterally. Notauli anterior parallel, parapsidal lines and parascutal carina absent. Scutellum weak alutaceous, around 0.8× as long as mesoscutum, broader than long, not overhanging metanotum, surface with some sparse short setae, slightly pointed distally; scutellar foveae absent; superficial, shiny anterior scutellar depression present. Mesopleuron and mesopleural triangle alutaceous (Figure 5f), almost without setae; axillula alutaceous (Figure 5c), with few sparse setae; dorsellum alutaceous, subrectangular, metapleural sulcus reaching mesopleuron at half of its height. Propodeum alutaceous, glabrous; medial carina present, with rugose carinae. Nucha short alutaceous to smooth.

Legs: Tarsal claws with short tooth. (Figure 5e).

Forewing: Hyaline, slightly longer than body. Coastal margin with cilia; Radial cell around four times as long as wide; Rs straight; areolet large; Rs + M incomplete, Cu1 separated (Figure 3a).

Metasoma: Shorter than head + mesosoma (Figure 3a), slightly longer than high in lateral view. Metasomal tergites without setae, smooth and shiny. Prominent part of ventral spine of hypopygium short, tapering to apex, around 1.4× as long as wide, with very few long sparse setae laterally that extend beyond apex of spine.

Gall: In staminate flowers of catkin inflorescences. Single-chambered (monothalamous), often in clusters, smooth centre with trichomes on the border, golden yellow colour and oblong/oval shape. No larger than 1.2 mm. (Figure 3b).

Males

Unknown.

Asexual Generation

Body length = 1.1 mm (n = 3).

Colour: Mostly brown, tarsi, scape, pedicel and the first two antennal segments light yellow. Structure and sculptures as the sexual generation (Figure 3c) with the following differences: 12 antennal segments, ocelli not raised and tarsal claw almost simple.

Gall: Cryptic gall in nodes of branches, often found adjacent to leaf scars and side branches, monothalamous, no longer than 2.8 mm (Figure 3d,e).

Etymology: The species' name refers to the location of the host tree on which it was first found: Outside of 'Valhalla', a graduate-student run bar on Rice University campus. The bar is named Valhalla, which is a great hall in Norse mythology where 'Odin receives the souls of heroes fallen bravely in battle' (O'donoghue, 2007).

Life history

Biology: Like most other known members of Cynipini, *Neuroterus valhalla* has two alternating generations that induce

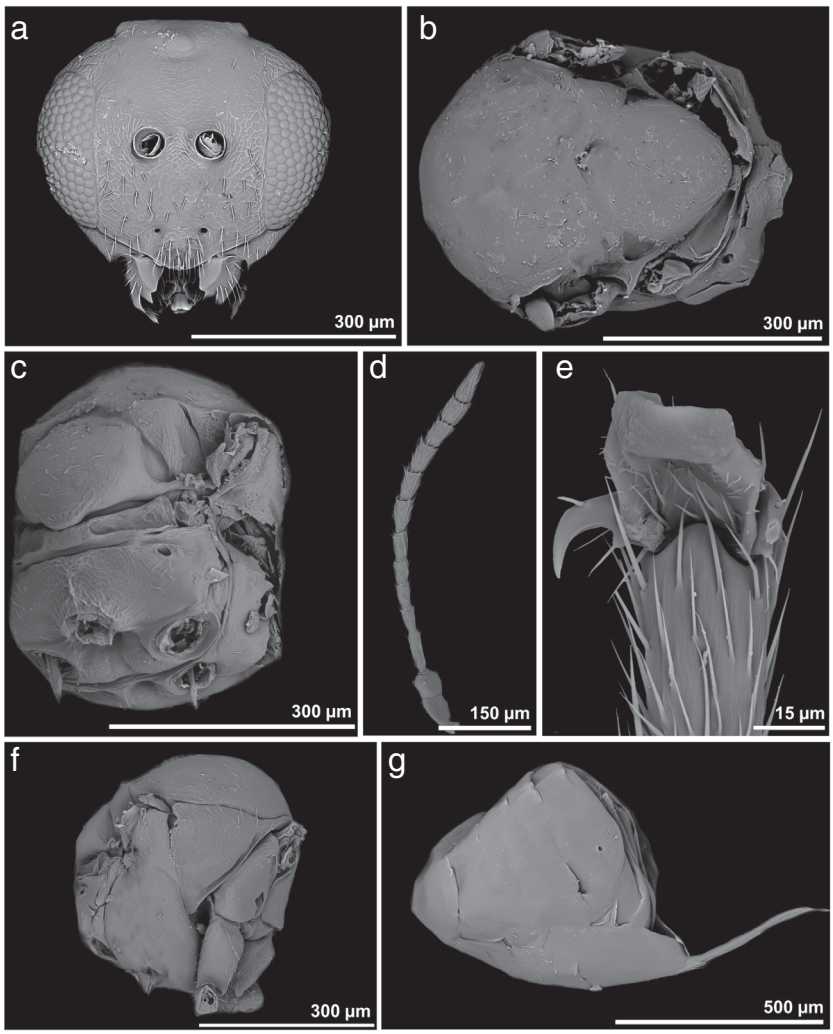

Fig. 5. Scanning electron microscopies of *N. valhalla* catkin generation female. (a) head anterior view. (b) mesosoma dorsal view; (c) mesosoma posterior view. (d) antennae. (e) tarsal claws. (f) mesosoma lateral view; (g) metasoma lateral view.

galls where their larvae feed (Figure 1). The stem node (asexual) generation wasp (Figures 1C and 3a) develops over the course of approximately 11 months within minute crypt galls found on the stem nodes (Figures 1E and 3f1), from which it will emerge in synchrony with the host's flowering phenology, typically during February (Figure 4). They will live as adults for approximately 2 days, as determined by lab rearing, only to find and oviposit in developing catkin buds (Figure 3b), where minute golden oval galls will develop (Figures 1B and 3c1). The catkin (sexual) generation wasps (Figures 1A and 3d) will then swiftly develop over the course of about 3 weeks within these gall structures and emerge in March (Figure 4). Upon emergence, adults live for approximately 2 days, possibly mate (if males are confirmed in the future), and oviposit into stem nodes near a leaf insertion (Figures 1D and 3e), completing the cycle.

Host: The population studied here uses *Q. virginiana* as its host. Additionally, we have incidentally collected two adult *N. valhalla* individuals emerging from *Quercus geminata* (Weinersmith et al., 2020), which were confirmed to be the same lineage (Figure 2).

Distribution: Confirmed in southeast Texas (Harris county) and the panhandle of Florida (Walton county). Most likely, extending to match much of the distribution of its host plants, *Q. geminata* and *Q. virginiana*, across the coastal southeastern United States, and potentially throughout the range of American live oaks in the subsection Virentes.

Genome description

The genome (GenBank accession WSXT00000000.1) has a GC content of 34.86%, and predicted size of 1.1 Mbp. This is a below-average genome size for the Cynipidae family (average 1.45 Mbp; Table 2), and the smallest genome reported to date

Table 2. Cynipidae genomes available in the NCBI database

Species	Genome size (Mbp)	Scaffold N50	Scaffold L50	Total Scaffolds	Tribe
Belonocnema kinseyi	1.538	150,973,230	5	5520	Cynipini
Andricus quercuslanigera	1.336	654,513	10	272,630	Cynipini
Synergus gifuensis	0.276	556,258	81	18,504	Synergini
Andricus grossulariae	1.412	440,296	13	229,755	Cynipini
Synergus itoensis	0.266	362,131	98	26,122	Synergini
Neuroterus valhalla	**1.117**	**344,082**	**14**	**334,575**	**Cynipini**
Synergus japonicus	0.226	61,479	941	12,796	Synergini
Synergus umbraculus	0.235	49,302	987	20,256	Synergini
Andricus inflator	1.869	1665	284,858	2,014,038	Cynipini
Neuroterus quercusbaccarum	2.569	1664	409,513	2,812,183	Cynipini
Andricus quercusramuli	1.937	1138	461,030	2,578,124	Cynipini
Andricus curvator	1.713	1116	402,730	2,458,281	Cynipini
Pseudoneuroterus saliens	2.06	970	536,641	3,378,461	Cynipini

Neuroterus valhalla is highlighted in bold.

within the tribe Cynipini (Table 2). The assembly has a scaffold N50 of 344 Kbp, L50 of 14, and an average coverage of 110×. A universal single-copy ortholog search (BUSCO) resulted in 94.5% of Eukaryotic genes present, 81.6% of which are complete.

A repetitive element annotation revealed that interspaced repeats represent 64.5% of the genome (Table 3), most of which are unclassified repetitive sequences (42.7%), and retrotransposons (10.94%). A preliminary *in-silico* gene annotation approach using only ab initio and protein-based annotation tools resulted in 32,005 predicted genes, including 18,802 multiexon protein-coding genes (Table S1).

In an ortholog gene group (orthogroup) analysis, 18 044 (56.4%) of the genes were assigned to 8186 orthogroups, 61 of which were species-specific (Table S2). As expected, the biggest assignment overlap was with *B. kinseyi*, which is the only other Cynipidae wasp with an annotated genome (13,561 genes in 7623 orthogroups, Tables S3 and S4), followed by *Nasonia vitripennis* (10,359 genes in 6428 orthogroups, Tables S3 and S4), which has the best annotated genome of the Chalcidoidea. The same analysis revealed that the genome of both *N. valhalla* and *B. kinseyi* has the two highest number of gene duplication events within genomes analysed (5411 and 11,461 respectively; Table S5). Given the low taxonomic resolution of the analysis, it is likely that most of these are associated with duplication events within Cynipidae.

Neuroterus valhalla within the Cynipini phylogeny

Both the COI phylogeny (Figure 2) and UCE phylogeny (Figure 6) have recovered the genus *Neuroterus* as polyphyletic, despite the different taxa represented in each data type. Yet, *N. valhalla* grouped with other Nearctic *Neuroterus* species in both datasets, while Palearctic *Neuroterus* species consistently grouped with representatives of other Cynipini genera such as

Table 3. Repetitive element composition in the *N. valhalla* genome

	Number of elements	Length occupied (Kbp)	Percentage of genome
Retrotransposons	335,625	122,082	10.93%
LINEs	288,091	102,889	9.21%
LTR elements	47,534	19,192	1.72%
Penelope	40,601	10,383	0.93%
SINEs	0	0	0.00%
DNA transposons	157,807	38,886	3.48%
Tc1-IS630-Pogo	62,442	16,000	1.43%
hobo-Activator	27,370	7497	0.67%
PiggyBac	3549	987	0.09%
Tourist	320	46	0.00%
Other	2988	604	0.05%
Helitron	202,382	82,541	7.39%
Unclassified	2,296,543	477,111	42.70%
Total interspersed repeats:		720,620	64.50%
Small RNA	20,793	2917	0.26%
Satellite	1582	287	0.03%
Simple repeats	401	14	0.00%

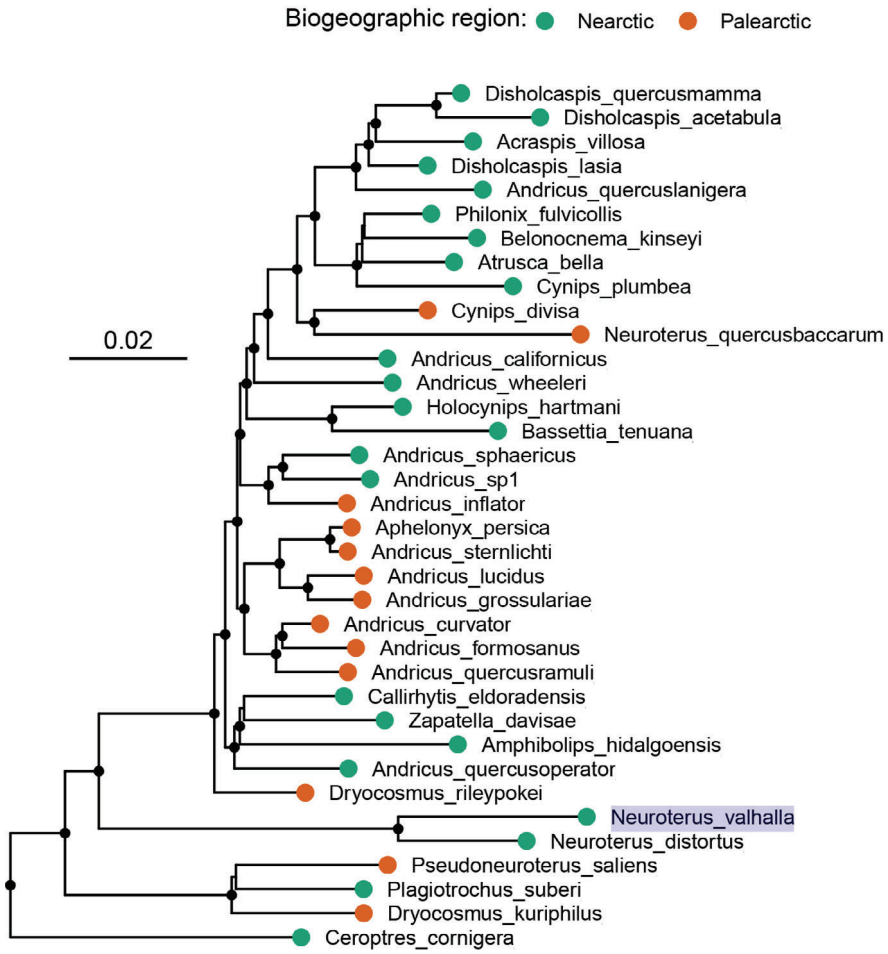

Fig. 6. Maximum-likelihood UCE phylogeny of Cynipini. Tree based on the 50% complete UCE matrix using best partitions selected by ModelFinder. Strongly supported nodes (≥80% SH-aLRT and ≥95% UFBoot values) are shown with black dots. Branch lengths represent genetic distance.

Cynips and *Andricus*. Additionally, every Cynipini genus represented by more than one species was recovered as polyphyletic. Those included *Andricus* Hartig, *Cynips* L., *Callirhytis* Foerster, *Disholcaspis* Dalla Torre & Keiffer and *Dryocosmus* Giraud. This taxonomic disagreement suggests that a larger, global revision of the genera within the Cynipini is needed.

Discussion

Neuroterus valhalla **sp. nov.** is described as the 15th known cynipid galler on the *Virentes* sections oaks (Table 1), which increases our knowledge and understanding of this diverse community. The species is also described from both generations simultaneously (Figure 1) linked by DNA barcoding, phenology, morphology and behavioural assays of oviposition preference, which is unusual for new species descriptions of Cynipini that often include a single generation (e.g. Melika & Abrahamson, 1997, 2000). In this study, we also build on the integrative species description format by incorporating a draft genome of the species, which, to our knowledge, is the first insect species to be described along with its sequenced genome. This also appears to be the second animal to have its genome published alongside its species description, the first being a close relative of the nematode *Caenorhabditis elegans* (Maupas, 1900) (Kanzaki et al., 2018). There has also been at least one other discovery, a 'new species' in the context of a phylogenomic analysis where its genome sequence is already known (Christmas et al., 2021). However, there was no formal species description in that case. The integration of a fully sequenced genome into a new species description remains rare in taxonomy; however, we hope that our work will pave the way for normalizing this practice, where applicable and where resources are available.

Integrating genomes with taxonomy

Sequencing the genome of a new species is an important step in understanding its biology, taxonomy, ecology and history. Advantages include (1) generating extensive genetic data that can be used for population genetic analyses and phylogenetic analysis of closely related species, (2) providing information

about how genes are distributed and arrayed across the genome and how the genome works as a whole, (3) serving as another complimentary source of information for species diagnostic characters, such as DNA sequence variation (e.g., DNA barcodes; Hebert et al., 2003), large insertions or deletions (Meusnier et al., 2004), chromosomal rearrangements (Stathos & Fishman, 2014), and other structural characteristics of a genome (Weissensteiner et al., 2020). In the current study, one major advantage of generating a draft genome was that these data were used to extract a UCE dataset to incorporate into a recent UCE phylogeny (Blaimer et al., 2020) to understand its evolutionary relationships within the group. Moreover, as more genomes become available, future comparative genomics studies across other cynipid species will improve our understanding of genome evolution in this species-rich group, including the observation of a high rate of gene duplications.

Another important advantage of the assignment of a reference genome at the moment of species description is related to the practical similarities between a reference genome and a type specimen. A reference genome has a relevance paralleling that of the type specimen: when that first genome is published for a species, it gets anchored to that species name, such that characteristics associated with that particular genome are therefore tied to that species. Providing both at once immediately connects morphological characters to genomic characters, all tied to a single species name. The parallel between the type specimen and the reference genome is not novel. Genbank itself has already incorporated a similar concept into its database (Federhen, 2015), and the discussion has come up in the context of a human genome reference (Ballouz et al., 2019). For instance, while morphology-focused taxonomic reviews of a group generally require authors to study the type specimens, we argue for the same reasons that taxonomic reviews employing genetic data should use DNA sequences extracted from the type genome (or reference genome) of a species.

In ideal cases, the same individual could be both the type specimen and the DNA donor for the sequencing of the type genome. For minute-sized organisms like *N. valhalla*, this will not always be possible, and a type genome will necessarily come from a destructively sampled syntype. We note that the desire to associate a reference genome with a holotype had been previously recognized, and there have been attempts to sequence genomes from type specimens (e.g. Boo et al., 2016; Hughey et al., 2014). However, specimen preservation methods often make DNA extraction, especially the high-quality extractions required for genomic DNA sequencing, highly challenging (but see Staats et al., 2013). The preference for fresh material is one more reason for new species descriptions to include a type genome, or to proactively consider preserving genetic material from the type or a syntype for the purpose of future genome sequencing in occasions where that is not possible.

Integrative taxonomy

While many species descriptions are unable to contain a fully sequenced genome or include complete life history information due to species rarity, lack of resources or other factors, we reinforce the need for more integrative species descriptions across cynipid systems. In particular, providing information about host association, DNA barcoding data (from multiple loci, if possible) and preserved material for genetic analysis, will allow better and more complete understanding of these complex systems. Although the genome sequence reported here was obtained from a single short-read sequencing run, it illustrates what is more likely to be available at lower costs for a broader range of taxonomists who may seek to add a sequenced genome to their species descriptions. However, as sequencing costs decrease, and more accurate and user-friendly assembly and annotation tools become available, we expect draft genome sequences of increasing quality (e.g., better scaffolding and annotation) to be associated with new species descriptions. Nonetheless, a draft genome assembly is enough for the purposes of a species description, as it readily provides the scientific community with genomic resources for the species and may be improved by future sequencing efforts.

Unseen dweller

Notably, these samples of *Neuroterus valhalla* were primarily found in the city of Houston, Texas (population > 2.3 Million), serving as an example of undescribed biodiversity hiding in a major and well-travelled urban centre – especially as this new gall wasp species itself harbours a diverse community of inquilines, parasitoids, hyperparasitoids and commensal associates. Previous experience with gall communities in this system suggest that up to 25% of insects may be unknown to science, given the occurrence of both generalists and specialists (Forbes et al., 2016; Weinersmith et al., 2020). For example, we have collected at least 12 morphospecies of natural enemies and/or associates that were raised from our stem and catkin collections (Figure S2). However, challenges in isolating the cryptic and minute gall structures here from adjacent and cryptic gall species and other woodborers make it difficult to establish trophic links in the current study. We report their associations here with the hope that future research can isolate and connect these unseen arthropods with their ecology. Nonetheless, many of the insects associated with this species or to other insects in the system are likely undescribed.

While we found at least 117 emerging *N. valhalla* specimens over the oak flowering seasons, we sampled approximately 330,000 flowers (as estimated by the weight of collected material), or approximately 8 L of catkin material. Given a rough estimation of the number of catkin flowers produced by a single tree, the known *Q. virginiana* population of 2270 trees on the 300-acre university campus (Arboretum, 2014), and the flower galling rate observed in our study (0.035%); we estimate that anywhere between 1×10^9 and 1×10^{10} *N. valhalla* individuals emerge every year on the campus population alone. While it may seem surprising that such a high number of insects would go unnoticed in a major urban centre, we note that *N. valhalla* galls are quite small and inconspicuous (Figure 3b,e), the emergent adult wasps are only 1.1–1.2 mm in size, and live for just a few

days following their emergence during a restricted annual temporal window (Figure 4). Additionally, we expect this number to fluctuate widely across years, given the variable nature of oak flowering (Sork, 1993).

Potentially unique life cycle

As described above, wasps in the tribe Cynipini have a bivoltine life cycle that alternates between a parthenogenetic generation comprised only of parthenogenic females, and a sexual generation comprised both males and females, which is known as heterogony (Figure 1; Folliot, 1964; Pujade-Villar et al., 2001). Although *N. valhalla* has two alternating generations, we did not observe a single male (n = 14 stem node generation individuals, n = 117 catkin generation individuals across 3 years), and we are therefore unable to confirm if any of the generations reproduce sexually. The most likely possibility here is that we have missed collecting males and/or there are some factors contributing to the rarity of males. The first possible factor is a skewed sex ratio, which could be caused by *Wolbachia* in some insects, however, *Wolbachia* is known to have no effect in the sex ratio of Cynipini specifically, despite high infection rates (Askew, 1960; Rokas et al., 2002). A second possibility would be a difference in phenology, with males emerging either earlier or later than females. However, we have maintained the galls and tissues over several months and continued to check emergence before and after the peak of season, with no signs of delayed or early male emergence. Finally, it is possible that the smaller size of males made them more prone to dehydration following gall removal from the host tree, resulting in them dying prior to emergence in the laboratory. Further studies would be needed to address any of these and other possibilities.

Notably, some of the most closely related species described to date in the subgenus *Diplobius*, including *N. fusifex* Pujade-Villar and *Neurocentropus vernus* Gill have their catkin-emerging generation as the sexual one (Patterson, 1928; Pujade-Villar et al., 2016). This is also true for *A. quercus-lanigera* (Ashmead 1981), which is the other described species galling catkin tissues in the same hosts (Hamel, 1973; Hood et al., 2018). Therefore, the catkin generation is more likely to be the sexual one in *N. valhalla*, despite the high number of individuals collected here (n = 117) in the absence of a male specimen. Interestingly, we observed individuals of *N. valhalla* from the catkin generation ovipositing into stem nodes directly after emergence without having had contact with males (Figures 1D and S1), which could be interpreted as evidence for parthenogenesis. This means that either both generations can reproduce parthenogenetically, or that, contrary to our original hypothesis, catkin-emerging insects represent the asexual generation, and that oviposited eggs would give rise to males of the stem node generation. This later alternative not only contradicts the trend among its most closely related species, but also the trend within Cynipini where the asexual generation is the longest lasting one (Zhang et al., 2019). For these reasons, we gravitate towards the first possibility where both generations can reproduce parthenogenetically, at least facultatively. Nonetheless, we cannot confirm this from the observations and data provided here alone. A population-wide study looking for the Hardy-Weinberg equilibrium of diallelic genetic markers (as in Stone et al., 2008), or an observational experiment confirming the completion of the life cycle in the absence of males is needed to verify this.

If the total absence of males is confirmed, however, it would imply that both of *N. valhalla*'s generations are fully parthenogenic. The presence of two alternating parthenogenic generations has never been described within Cynipidae before, rendering *N. valhalla* a possible unique case. Currently, only five exceptions to heterogony (alternating sexual and asexual generations) are confirmed in the group. Three of these (*Dryocosmus kuriphilus*, *Aandricus targionii* and *A. pseudoflos*) skipped the sexual generation altogether and swapped to having a life cycle with a single parthenogenetic generation (Buffington et al., 2020), whereas *D. zhulli* contrastingly skipped the asexual form (Zhu et al., 2015), and *Aradus quadrilineatus* can skip either generation (Folliot, 1964). However, many other species are only known by one of the two generations, likely due to incomplete knowledge of their life cycle (Stone et al., 2008).

Hosts and distribution

The *N. valhalla* population primarily described here uses the southern live oak (*Q. virginiana*) as its host. However, in a previous study we have also collected two *N. valhalla* individuals emerging from *Q. geminata* stems collected in northern Florida, USA (Weinersmith et al., 2020). One of these individuals showed a mitochondrial haplotype intermediate to those found in the *Q. virginiana* host in Texas, hinting at a mitochondrial connectivity across the U.S. Gulf coast. *Quercus virginiana* and *Q. geminata* are closely related oak species (Cavender-Bares & Pahlich, 2009; Manos & Hipp, 2021) with an extensive known overlap of Cynipidae gall wasp species comprising a complex system (Table 1; Egan et al., 2013; Price et al., 2004; Zhang et al., 2019). However, there is a frequent break in the distribution of gall species or genetic structure within species in this system around the Mississippi river (Zhang et al., 2021), so finding the species beyond this break and using an alternative host suggests that its true range covers the distribution of its host species, from central Texas passing through the U.S. Gulf coast to Florida and North Carolina (Cavender-Bares & Pahlich, 2009).

Genome features

The genome of *N. valhalla* has a predicted size of 1.1 Mbp, the smallest genome size reported to date within the Cynipidae tribe Cynipini (Table 2). This observation could be associated with the organism's quick developmental rate in the catkin generation, as it has been previously suggested that genome size in arthropods can be negatively correlated with developmental rate (Gregory, 2002). Also, genome size reduction can occur as a consequence of miniaturization of body size, as genome

size is directly correlated with cell size, and the latter can be a constraint for minute insects (Hanken & Wake, 1993; Lindsey et al., 2018; Polilov, 2015). Therefore, both quick developmental rate and minute size might contribute to this relatively small genome size. However, there is a big phylogenetic component to genome size in this case, given that Cynipini as a whole have much larger genomes than most Hymenoptera. While the average Cynipini genome size is 1.72 Mbp (n = 9; Table 2), hymenopteran genomes range mostly between 0.18 and 0.34 Mbp (Branstetter et al., 2018; Hotaling et al., 2021). Therefore, while *N. valhalla*'s genome is small relative to other Cynipini genomes, it is still large relative to the rest of Hymenoptera. Interestingly, other cynipids in the sister tribe Synergini have shorter genomes (0.25 Mbp on average, hence 4.5× smaller than that of *N. valhalla*; Table 2), following the more general trend in Hymenoptera. By using the topology recovered by Blaimer *et al.* (2020), where Synergini is older and gives rise to Cynipini, we infer that the increased genome size is a synapomorphy of the latter. Under this observation, we propose five alternative hypotheses/ mechanisms for the increased genome size in Cynipini.

The first (1) is the null hypothesis, which is that this genome size increase was caused by the expansion of repetitive elements, without meaningful evolutionary or ecological cause beyond a selfish nature of repetitive elements (Orgel & Crick, 1980). However, there are a couple of notable life history differences between Synergini and Cynipini that may suggest otherwise. While the latter can produce galls from undifferentiated meristematic tissue, Synergini are mostly inquilines that lay their eggs in the galls of other gall-forming species, inducing only the nutritive tissue they consume (Sanver & Hawkins, 2000). There is evidence that parasitism poses constraints to genome size in Hymenoptera (Ardila-Garcia et al., 2010). In this case, we may hypothesize (2) that Cynipini's move towards galling and away from 'traditional' parasitism may have released this evolutionary constraint, resulting in a genome expansion in the group, again possibly by the expansion of repetitive elements. The third (3) hypothesis suggests that the evolution of gall production may be associated with an extensive gene diversification, which may be needed for the manipulation of host plants, and consequent genome expansion process. However, recent comparative analysis of *Synergus itoensis* Abe, Ide & Wachi genome, an inquiline that has independently gained the ability to induce galls, with other inquiline *Synergus* species suggests that gene duplication may not be associated with gall-forming capacity within Cynipidae (Gobbo et al., 2020).

The other notable difference between the two tribes is Cynipini's unique life cycle discussed above – heterogony (Figure 1). In light of this observation, the fourth (4) hypothesis for the larger Cynipini genome is that the adoption of a (partial) parthenogenetic life history may have caused genome expansion, given that asexual reproduction has deep genetic consequences. For example, it can lead to the activation of transposable elements (Hickey, 1982) or accumulation of palindromes to escape Muller's ratchet (Jaron et al., 2019; Muller, 1932; Rozen et al., 2003). However, a review found no evidence that genome size increases in response to parthenogenesis (Jaron et al., 2019). Lastly, the fifth (5) hypothesis revolves around this unique life cycle, which implies that a single genome must generate what are eco-functionally two species, including two morphologically distinct adults with distinct behaviours, which induce galls in different tissues. This phenomenon involves either a highly complex gene regulation system, extensive pleiotropy or as suggested by our orthogroup analysis, gene duplication and genome expansion (Table S5). Therefore, it is possible that this complex regulatory process involves a genome size increase and may be a major contributor to Cynipini's expanded genomes. While we see this hypothesis as the most likely among the five proposed here, further evidence is needed to better evaluate this relationship. For example, a comparative transcriptome analysis across two generations within the same species, or an evaluation of the genome size of Pediaspidini, the *Acer* sp. gallers that independently evolved heterogony (Pujade-Villar et al., 2001), could shed much light onto this.

Our genome annotation resulted in a high number of genes (32005), but only a portion of those (18 044, 56.4%) could be assigned to an orthogroup (Table S2). The raw number of predicted genes can be dependent on the annotation pipeline (Branstetter et al., 2018; Elsik et al., 2014), so it is possible that our annotation overestimated the true number of genes. Yet, the only other annotated Cynipidae genome (*Belonocnema kinseyi*, GenBank accession GCA_010883055.1), which has a high-quality chromosome level assembly and evidence-based annotation, has 25,246 total genes, which is also relatively high within insect genomes (Rosenfeld et al., 2016). This high number of genes within the group may be associated with gene duplications, as suggested by the orthogroup analysis (Table S5), which we hypothesize is associated with Cynipini's life cycle as discussed above. Nonetheless, further analysis of better characterized Cynipidae genomes is needed to conclude on the causes and mechanisms behind both the increased genome size and high gene count.

Finally, we should note that due to their haplodiploidy, hymenopteran genome sequencing libraries can be built with a single haploid male, which has the potential to greatly improve the genome assembly quality (Hearn et al., 2014; Yahav & Privman, 2019). However, we were unable to find the males for this particular species, and a single female individual was used instead.

Phylogenetic placement

Both of the phylogenetic reconstructions here (Figures 2 and 6) independently recovered *Neuroterus* Hartig as a polyphyletic genus, which had been observed multiple times in the past literature (e.g. Blaimer et al., 2020; Melika et al., 2010). The genus was originally described from a Palearctic lineage (Hartig, 1840), and now includes ∼90 described species, more than half of which are known from North America (Burks, 1979; Melika et al., 2010). *Neuroterus* has been regarded as a morphologically challenging one since its origin (Kinsey, 1923), mainly due to the minute size of the insects. *Neuroterus* are among the smallest of the Cynipidae, and their galls are usually

more 'primitive', lacking the highly specialized tissues and layers characteristic of other cynipid genera (Melika, 2006). Its defining feature, the lack of the transscutal articulation, is quite common among small-sized cynipids, which resulted in the grouping of several independent lineages into the same genus (Pujade-Villar et al., 2014).

Therefore, *Neuroterus* is likely an example of convergent miniaturization-related morphological features being used to define a group resulting in a polyphyletic arrangement, a common occurrence among miniaturized taxa (Rundell & Leander, 2010). With that said, a revision of the genus *Neuroterus* is far beyond the scope of our work, and we understand that *N. valhalla*, along with other members of the subgenus *Diplobius*, will likely be relocated to a different genus in the future, with *Neuroterus* being restricted to the originally described Palearctic lineage (Hartig, 1840; Pujade-Villar et al., 2014). Furthermore, every other genus represented by more than one species in our phylogenies was recovered as polyphyletic, suggesting that Cynipini as a whole need a major systematic review.

Conclusion

In this current study, we describe a new species of gall wasp associated with American live oaks (*Virentes*), *Neuroterus valhalla* **sp. nov.**, along with its complete life cycle and a draft genome sequence. By integrating a draft genome with the species description, we highlight the parallel between the type specimen and the reference genome, and hope to open the path to what we believe is the next step in integrative taxonomy. Moreover, we discuss the possible reasons for enlarged genome size observed in Cynipidae, and further showcase the versatility of the genome data by extracting UCE loci and incorporating these into previous phylogenetic data. With this, we confirm previous studies that found *Neuroterus* to be polyphyletic, and suggest that the Nearctic members of the genus should be moved after a much needed taxonomic revision. Finally, this study also serves as an example for unexplored biodiversity present even in densely populated urban areas, hinting at fascinating biology that awaits future discovery.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix S1. Supporting Information

Acknowledgements

Daniel Sazer and Jordan Miller for the stacked picture system access; Lin-Yi Zhang for the help with genome scaffolding attempts with long-read data; Ellen Martinson for advice on genome annotation; Luisa Rezende for the assistance with image finishing; Michael Gates for access and assist with scanning electron microscopy and identification of parasitoids based on photos; Egan Lab members for the feedback and assistance throughout the project; Rice University and National Museum of Natural History IT professionals for assistance with computational needs for the project; and illustrator Barbara Rossi for drawing seen in Figure 1. Mention of trade names or commercial products in this publication is solely to provide specific information and does not imply recommendation or endorsement by the USDA. USDA is an equal opportunity provider and employer.

Conflict of interest

The authors declare no competing interests.

Data availability statement

Further information that supports the findings of this study is available in the supplementary material of this article, and data that support the findings of this study are openly available in Dryad at https://doi.org/10.5061/dryad.zgmsbccc8.

References

Abrahamson, W.G., Melika, G., Scrafford, R. & Csóka, G. (1998) Host plant associations and specificity among cynipid gall-inducing wasps of eastern USA. *The biology of gall-inducing arthropods*, pp. 226–240. USDA Forest Service, North Central Experiment Station. General Technical Report North Carolina-199, St. Paul, MN.

Ardila-Garcia, A.M., Umphrey, G.J. & Gregory, T.R. (2010) An expansion of the genome size dataset for the insect order Hymenoptera, with a first test of parasitism and eusociality as possible constraints. *Insect Molecular Biology*, **19**, 337–346.

Ashmead, W.H. (1881) On the cynipidous galls of Florida. *Transactions of the American Entomological Society*, **9**, 9–15.

Ashmead, W.H. (1885) On the cynipidous galls of Florida with descriptions of new species. *Transactions of the American Entomological Society*, **12**, 5–9.

Ashmead, W.H. (1896) Descriptions of new cynipidous galls and gall wasps in the United States National Museum. *Proceedings of the United States National Museum*, **19**, 113–136.

Askew, R.R. (1960) Some observations on *Diplolepis rosae* (L.)(Hym., Cynipidae) and its parasites. *Entomologist's Monthly Magazine*, **95**, 191–192.

Ballouz, S., Dobin, A. & Gillis, J.A. (2019) Is it time to change the reference genome? *Genome Biology*, **20**, 1–9.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S. *et al.* (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, **19**, 455–477.

Bányai, K., Kemenesi, G., Budinski, I., Földes, F., Zana, B., Marton, S. *et al.* (2017) Candidate new rotavirus species in Schreiber's bats, Serbia. *Infection, Genetics and Evolution*, **48**, 19–26.

Bao, Z. & Eddy, S.R. (2002) Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Research*, **12**, 1269–1276.

Bao, W., Kojima, K.K. & Kohany, O. (2015) Repbase update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*, **6**, 1–6.

Bassett, H.F. (1881) New species of Cynipidae. *The Canadian Entomologist*, **13**, 51–57.

Blaimer, B.B., Gotzek, D., Brady, S.G. & Buffington, M.L. (2020) Comprehensive phylogenomic analyses re-write the evolution of parasitism within cynipoid wasps. *BMC Evolutionary Biology*, **20**, 155.

Bolger, A.M., Lohse, M. & Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**(15), 2114–2120.

Boo, G.H., Hughey, J.R., Miller, K.A. & Boo, S.M. (2016) Mitogenomes from type specimens, a genotyping tool for morphologically simple species: ten genomes of agar-producing red algae. *Scientific Reports*, **6**, 1–13.

Borowiec, M. (2019) Spruceup: fast and flexible identification, visualization, and removal of outliers from large multiple sequence alignments. *Journal of Open Source Software*, **4**, 1635.

Branstetter, M.G., Longino, J.T., Ward, P.S., Faircloth, B.C. & Price, S. (2017) Enriching the ant tree of life: enhanced UCE bait set for genome-scale phylogenetics of ants and other Hymenoptera. *Methods in Ecology and Evolution*, **8**, 768–776.

Branstetter, M.G., Childers, A.K., Cox-Foster, D., Hopper, K.R., Kapheim, K.M., Toth, A.L. & Worley, K.C. (2018) Genomes of the Hymenoptera. *Current Opinion in Insect Science*, **25**, 65–75.

Buermans, H.P.J. & Den Dunnen, J.T. (2014) Next generation sequencing technology: advances and applications. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, **1842**, 1932–1941.

Buffington, M.L., Forshage, M., Liljeblad, J., Tang, C.T. & van Noort, S. (2020) World Cynipoidea (Hymenoptera): a key to higher-level groups. *Insect Systematics and Diversity*, **4**, 1.

Buhr, H. (1965) *Bestimmungstabellen der gallen (zoo- und phytocecidien) an pflanzen mittel-und nordeuropas*. Fischer Verlag, Jena.

Burks, B. (1979) Superfamily Cynipoidea. *Catalog of Hymenoptera in America North of Mexico* (ed. by K.V. Krombein, P. Hurd, D.R. Smith and B. Burks), pp. 1045–1107. Smithsonian Institution Press, Washington, DC.

Caballero, M. & Wegrzyn, J. (2019) gFACs: gene filtering, analysis, and conversion to unify genome annotations across alignment and gene prediction frameworks. *Genomics, Proteomics & Bioinformatics*, **17**, 305–310.

Castresana, J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, **17**, 540–552.

Cavender-Bares, J. & Pahlich, A. (2009) Molecular, morphological, and ecological niche differentiation of sympatric sister oak species, *Quercus virginiana* and *Q. geminata* (Fagaceae). *American Journal of Botany*, **96**, 1690–1702.

Cavender-Bares, J., González-Rodríguez, A., Eaton, D.A., Hipp, A.A., Beulke, A. & Manos, P.S. (2015) Phylogeny and biogeography of the American live oaks (Quercus subsection *Virentes*): a genomic and population genetics approach. *Molecular Ecology*, **24**, 3668–3687.

Christmas, M.J., Jones, J.C., Olsson, A., Wallerman, O., Bunikis, I., Kierczak, M. *et al.* (2021) Genetic barriers to historical gene flow between cryptic species of alpine bumblebees revealed by comparative population genomics. *Molecular Biology and Evolution*, msab086. **38**(8), 3126–3143.

Cook, J.M., Rokas, A., Pagel, M. & Stone, G.N. (2002) Evolutionary shifts between host oak sections and host-plant organs in *Andricus* gallwasps. *Evolution*, **56**, 1821–1830.

Cook, L.G., Edwards, R.D., Crisp, M.D. & Hardy, N.B. (2010) Need morphology always be required for new species descriptions? *Invertebrate Systematics*, **24**, 322–326.

Costello, M.J., Wilson, S. & Houlding, B. (2012) Predicting total global species richness using rates of species description and estimates of taxonomic effort. *Systematic Biology*, **61**, 871.

Costello, M.J., May, R.M. & Stork, N.E. (2013) Can we name Earth's species before they go extinct? *Science*, **339**, 413–416.

Cozzuol, M.A., Clozato, C.L., Holanda, E.C., Rodrigues, F.H., Nienow, S., De Thoisy, B. *et al.* (2013) A new species of tapir from the Amazon. *Journal of Mammalogy*, **94**, 1331–1345.

Dayrat, B. (2005) Towards integrative taxonomy. *Biological Journal of the Linnean Society*, **85**, 407–417.

Deans, A.R., Yoder, M.J. & Balhoff, J.P. (2012) Time to change how we describe biodiversity. *Trends in Ecology and Evolution*, **27**, 78–84.

Dirzo, R., Young, H.S., Galetti, M., Ceballos, G., Isaac, N.J. & Collen, B. (2014) Defaunation in the Anthropocene. *Science*, **345**, 401–406.

Duran, D.P., Herrmann, D.P., Roman, S.J., Gwiazdowski, R.A., Drummond, J.A., Hood, G.R. & Egan, S.P. (2019) Cryptic diversity in the north American Dromochorus tiger beetles (Coleoptera: Carabidae: Cicindelinae): a congruence-based method for species discovery. *Zoological Journal of the Linnean Society*, **186**, 250–285.

Duran, D.P., Laroche, R.A., Gough, H.M., Gwiazdowski, R.A., Knisley, C.B., Herrmann, D.P. *et al.* (2020) Geographic life history differences predict genomic divergence better than mitochondrial barcodes or phenotype. *Genes*, **11**, 265.

Ebach, M.C., Valdecasas, A.G. & Wheeler, Q.D. (2011) Impediments to taxonomy and users of taxonomy: accessibility and impact evaluation. *Cladistics*, **27**, 550–557.

Egan, S.P., Hood, G.R., DeVela, G. & Ott, J.R. (2013) Parallel patterns of morphological and behavioral variation among host-associated populations of two gall wasp species. *PLoS One*, **8**, e54690.

Egan, S.P., Weinersmith, K.L., Liu, S., Ridenbaugh, R.D., Zhang, Y.M. & Forbes, A.A. (2017) Description of a new species of Euderus Haliday from the southeastern United States (Hymenoptera, Chalcidoidea, Eulophidae): the crypt-keeper wasp. *ZooKeys*, **645**, 37.

Egan, S.P., Hood, G.R., Martinson, E.O. & Ott, J.R. (2018) Cynipid gall wasps. *Current Biology*, **28**, 1370–1374.

Elsik, C.G., Worley, K.C., Bennett, A.K., Beye, M., Camara, F., Childers, C.P. *et al.* (2014) Finding the missing honey bee genes: lessons learned from a genome upgrade. *BMC Genomics*, **15**, 1–29.

Emms, D.M. & Kelly, S. (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, **20**, 1–14.

Espirito-Santo, M.M. & Fernandes, G.W. (2007) How many species of gall inducing insects are there on earth and where are they? *Annuals of the Entomoogical Society of America*, **100**, 95–99.

Faircloth, B.C. (2015) PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics*, **32**, 786–788.

Federhen, S. (2015) Type material in the NCBI taxonomy database. *Nucleic Acids Research*, **43**, D1086–D1098.

Folliot, R. (1964) Contribution a l'étude de la biologie des cynipides gallicoles (Hymenopteres, Cynipoidea). *Annales des Sciences Naturelles*, **12**, 407–564.

Forbes, A.A., Hall, M.C., Lund, J., Hood, G.R., Izen, R., Egan, S.P. & Ott, J.R. (2016) Parasitoids, hyperparasitoids, and inquilines associated with the sexual and asexual generations of the gall former, *Belonocnema treatae* (Hymenoptera: Cynipidae). *Annals of the Entomological Society of America*, **109**, 49–63.

Forbes, A.A., Bagley, R.K., Beer, M.A., Hippee, A.C. & Widmayer, H.A. (2018) Quantifying the unquantifiable: why Hymenoptera, not Coleoptera, is the most speciose animal order. *BMC Ecology*, **18**, 1–11.

Frankie, G.W., Morgan, D.L., Gaylor, M.J., Benskin, J.G., Clark, W.E., Reed, H.C. & Hamman, P.J. (1977) The mealy oak gall [*Disholcaspis cinerosa*] on ornamental live oak [*Quercus virginiana*] in Texas. *MP-Texas Agricultural Experiment Station (USA)*.

Giam, X., Scheffers, B.R., Sodhi, N.S., Wilcove, D.S., Ceballos, G. & Ehrlich, P.R. (2012) Reservoirs of richness: least disturbed tropical forests are centres of undescribed species diversity. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 67–76.

Gobbo, E., Lartillot, N., Hearn, J., Stone, G.N., Abe, Y., Wheat, C.W. *et al.* (2020) From inquilines to gall inducers: genomic signature of

a life-style transition in Synergus gall wasps. *Genome Biology and Evolution*, **12**, 2060–2073.
Godfray, H.C.J. (2002) Challenges for taxonomy. *Nature*, **417**, 17–19.
Goodwin, S., McPherson, J.D. & McCombie, W.R. (2016) Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, **17**, 333.
Gregory, T.R. (2002) Genome size and developmental complexity. *Genetica*, **115**, 131–146.
Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, **59**, 307–321.
Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, **29**, 1072–1075.
Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J. *et al.* (2008) Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biology*, **9**, 1–22.
Hamel, D.R. (1973) *The biology and ecology of the live oak woolly leaf gall Andricus Laniger Ashmead (Hymenoptera: Cynipidae)*. Doctoral dissertation, Texas A&M University, College Station.
Hanken, J. & Wake, D.B. (1993) Miniaturization of body size: organismal consequences and evolutionary significance. *Annual Review of Ecology and Systematics*, **24**, 501–519.
Harris, R.A. (1979) Glossary of surface sculpturing. *Occasional Papers in Entomology*, **28**, 1–31.
Hartig, T. (1840) Ueber die Familie der Gallwespen. III. *Zeitschrift für Entomologie (Germar)*, **2**, 176–209.
Hartop, E.A., Brown, B.V. & Disney, R.H.L. (2015) Opportunity in our ignorance: urban biodiversity study reveals 30 new species and one new Nearctic record for Megaselia (Diptera: Phoridae) in Los Angeles (California, USA). *Zootaxa*, **3941**, 451–484.
Hassemer, G., Bruun-Lund, S., Shipunov, A.B., Briggs, B.G., Meudt, H.M. & Rønsted, N. (2019) The application of high-throughput sequencing for taxonomy: the case of Plantago subg. Plantago (Plantaginaceae). *Molecular Phylogenetics and Evolution*, **138**, 156–173.
Hearn, J., Stone, G.N., Bunnefeld, L., Nicholls, J.A., Barton, N.H. & Lohse, K. (2014) Likelihood-based inference of population history from low-coverage de novo genome assemblies. *Molecular Ecology*, **23**, 198–211.
Hebert, P.D. & Gregory, T.R. (2005) The promise of DNA barcoding for taxonomy. *Systematic Biology*, **54**, 852–859.
Hebert, P.D., Cywinska, A., Ball, S.L. & DeWaard, J.R. (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **270**, 313–321.
Hebert, P.D., Penton, E.H., Burns, J.M., Janzen, D.H. & Hallwachs, W. (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences*, **101**, 14812–14817.
Heraty, J. & Hawks, D. (1998) Hexamethyldisilazane: a chemical alternative for drying insects. *Entomological News*, **109**, 369–374.
Hickey, D.A. (1982) Selfish DNA: a sexually-transmitted nuclear parasite. *Genetics*, **101**, 519–531.
Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q. & Le, S.V. (2017) UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, **35**, 518–522.
Hood, G.R., Zhang, L., Topper, L., Brandão-Dias, P.F.P., Del Pino, G.A., Comerford, M.S. & Egan, S.P. (2018) 'Closing the life cycle' of *Andricus quercuslanigera* (Hymenoptera: Cynipidae). *Annals of the Entomological Society of America*, **111**, 103–113.
Hotaling, S., Sproul, J.S., Heckenhauer, J., Powell, A., Larracuente, A.L., Pauls, S.U. et al. (2021) Long-reads are revolutionizing 20 years of insect genome sequencing. *bioRxiv*, 2021.02.14.431146.
Hughey, J.R., Gabrielson, P.W., Rohmer, L., Tortolani, J., Silva, M., Miller, K.A. *et al.* (2014) Minimally destructive sampling of type specimens of Pyropia (Bangiales, Rhodophyta) recovers complete plastid and mitochondrial genomes. *Scientific Reports*, **4**, 1–10.
i5K Consortium (2013) The i5K initiative: advancing arthropod genomics for knowledge, human health, agriculture, and the environment. *Journal of Heredity*, **104**, 595–600.
International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
Jaron, K.S., Bast, J., Nowell, R.W., Ranallo-Benavidez, T.R., Robinson-Rechavi, M. & Schwander, T. (2019) Genomic features of asexual animals. *BioRxiv*, p497495.
Jatnika, M.F., Weerakoon, G., Arachchige, O., Noer, I.S., Voytsekhovich, A. & Luecking, R. (2019) Discoveries through social media and in your own backyard: two new species of *Allographa* (Graphidaceae) with pigmented lirellae from the Palaeotropics, with a world key to species of this group. *The Lichenologist*, **51**, 227–233.
Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K., von Haeseler, A. & Jermiin, L.S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, **14**, 587.
Kanzaki, N., Tsai, I.J., Tanaka, R., Hunt, V.L., Liu, D., Tsuyama, K. *et al.* (2018) Biology and genome of a newly discovered sibling species of *Caenorhabditis elegans*. *Nature Communications*, **9**, 1–12.
Katoh, K. & Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, **30**, 772–780.
Kieffer, J.J. (1910) Description de nouveaux Hyménoptères. Bolletino del Laboratorio di Zoologia Generale e Agraria della R. *Sciola superiore d'Agricola di Portici*, **4**, 105–117.
Kinsey, A.C. (1923) The gall wasp genus Neuroterus (Hymenoptera). *Indiana University Studies*, **58**, 1–150.
Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, **35**, 1547–1549.
Lewin, H.A., Robinson, G.E., Kress, W.J., Baker, W.J., Coddington, J., Crandall, K.A. *et al.* (2018) Earth BioGenome project: sequencing life for the future of life. *Proceedings of the National Academy of Sciences*, **115**, 4325–4333.
Liljeblad, J. & Ronquist, F. (1998) A phylogenetic analysis of higher-level gall wasp relationships (Hymenoptera: Cynipidae). *Systematic Entomology*, **23**, 229–252.
Lindsey, A.R., Kelkar, Y.D., Wu, X., Sun, D., Martinson, E.O., Yan, Z. *et al.* (2018) Comparative genomics of the miniature wasp and pest control agent *Trichogramma pretiosum*. *BMC Biology*, **16**, 1–20.
Lorch, J.M., Palmer, J.M., Vanderwolf, K.J., Schmidt, K.Z., Verant, M.L., Weller, T.J. & Blehert, D.S. (2018) Malassezia vespertilionis sp. nov.: a new cold-tolerant species of yeast isolated from bats. *Persoonia: Molecular Phylogeny and Evolution of Fungi*, **41**, 56.
Lynn R. Laurey Arboretum. (2014) *List of woody species on rice campus*. Available at: https://cpb-us-e1.wpmucdn.com/blogs.rice.edu/dist/8/3254/files/2014/06/List-of-Woody-Species-on-Campus.pdf [Accessed on 26th March 2021].
Maia, L.F., Nascimento, A.R. & Faria, L.D. (2018) Four years host–parasitoid food web: testing sampling effort on trophic levels. *Studies on Neotropical Fauna and Environment*, **53**, 132–142.
Manos, P.S. & Hipp, A.L. (2021) An updated Infrageneric classification of the north American oaks (*Quercus* subgenus *Quercus*): review of the contribution of Phylogenomic data to biogeography and species diversity. *Forests*, **12**, 786.
Mayr, G.L. (1881) Die Genera der gallenbewohnenden Cynipiden. *Jahresbericht der Rossauer Communal-Oberrealschule*, **20**, 1–38.
Medianero, E. & Nieves-Aldrey, J. (2017) First record of the oak gall wasp genus Neuroterus Hartig, 1840 (Hymenoptera, Cynipidae,

Cynipini) from Central America with description of three new species from Panama and Costa Rica. *Consejo Superior de Investigaciones Científicas (España)*, **73**, e057.

Melika, G. (2006) Gall wasps of Ukraine. Vols. 1 & 2, *Vestnik Zoologii*. *Journal of Schmalhausen Institute of Zoology Supplement*, **21**, 1–644.

Melika, G. & Abrahamson, W.G. (1997) Descriptions of four new species of cynipid gall wasps of the genus Neuroterus Hartig (Hymenoptera Cynipidae) with redescriptions of some known species from the eastern United States. *Proceedings of the Entomological Society of Washington*, **99**, 560–573.

Melika, G. & Abrahamson, W.G. (2000) Review of the cynipid gall wasps of the genus Locaulus Mayr (Hymenoptera: Cynipidae) with descriptions. *Proceedings of the Entomological Society of Washington*, **102**, 198–211.

Melika, G., Pujade-Villar, J., Abe, Y., Tang, C., Nicholls, J., Wachi, N. *et al.* (2010) Palaearctic oak gallwasps galling oaks (Quercus) in the section Cerris: re-appraisal of generic limits, with descriptions of new genera and species (Hymenoptera: Cynipidae: Cynipini). *Zootaxa*, **2470**, 1–79.

Melika, G., Buss, E.A., Nicholls, J.A., Bird, J.P. & Stone, G.N. (2013) Life cycle of *Disholcaspis quercusvirens* (Hymenoptera: Cynipidae) with a description of the sexual generation. *Florida Entomologist*, **96**, 991–1001.

Meusnier, I., Valero, M., Olsen, J.L. & Stam, W.T. (2004) Analysis of rDNA ITS1 indels in *Caulerpa taxifolia* (Chlorophyta) supports a derived, incipient species status for the invasive strain. *European Journal of Phycology*, **39**, 83–92.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., Von Haeseler, A. & Lanfear, R. (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, **37**, 1530–1534.

Muller, H.J. (1932) Some genetic aspects of sex. *American Naturalist*, **66**, 118–138. https://doi.org/10.1086/280418.

Muller, C.H. (1961) The live oaks of the series *Virentes*. *American Midland Naturalist*, **65**(1), 17–39.

Nicholls, J.A., Challis, R.J., Mutun, S. & Stone, G.N. (2012) Mitochondrial barcodes are diagnostic of shared refugia but not species in hybridizing oak gallwasps. *Molecular Ecology*, **21**, 4051–4062.

O'Donoghue, H. (2007) *From asgard to valhalla: the remarkable history of the norse myths*. Bloomsbury Publishing, London.

Orgel, L.E. & Crick, F.H. (1980) Selfish DNA: the ultimate parasite. *Nature*, **284**, 604–607.

Orr, M.C., Ferrari, R.R., Hughes, A.C., Chen, J., Ascher, J.S., Yan, Y.H. *et al.* (2021) Taxonomy must engage with new technologies and evolve to face future challenges. *Nature Ecology & Evolution*, **5**, 3–4.

Ou, S. & Jiang, N. (2018) LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiology*, **176**, 1410–1422.

Padial, J.M. & De la Riva, I. (2007) Integrative taxonomists should use and produce DNA barcodes. *Zootaxa*, **1586**, 67–68.

Padial, J.M., Miralles, A., De la Riva, I. & Vences, M. (2010) The integrative future of taxonomy. *Frontiers in Zoology*, **7**, 1–14.

Patterson, J.T. (1928) Sexes in the Cynipidae and male-producing and female-producing lines. *The Biological Bulletin*, **54**, 201–211.

Pearson, D.L., Hamilton, A.L. & Erwin, T.L. (2011) Recovery plan for the endangered taxonomy profession. *Bioscience*, **61**, 58–63.

Penzes, Z., Tang, C.T., Stone, G.N., Nicholls, J.A., Schweger, S., Bozso, M. & Melika, G. (2018) Current status of the oak gallwasp (Hymenoptera: Cynipidae: Cynipini) fauna of the eastern Palaearctic and oriental regions. *Zootaxa*, **4433**, 245–289.

Polilov, A.A. (2015) Small is beautiful: features of the smallest insects and limits to miniaturization. *Annual Review of Entomology*, **60**, 103–121.

Price, P.W., Abrahamson, W.G., Hunter, M.D. & Melika, G. (2004) Using gall wasps on oaks to test broad ecological concepts. *Conservation Biology*, **18**, 1405–1416.

Price, A.L., Jones, N.C. & Pevzner, P.A. (2005) De novo identification of repeat families in large genomes. *Bioinformatics*, **21**, i351–i358.

Pujade-Villar, J., Melika, G., Bellido, D. & López, G.S. (2001) Current state of knowledge of heterogony in Cynipidae (Hymenoptera, Cynipoidea). *Sessió Conjunta d'Entomologia*, **11**, 87–107.

Pujade-Villar, J., Tovar, D.C., Ruíz, U.M.B. & Melika, G. (2014) First record of *Neuroterus* galls on twigs in Mexico with description of two new species (Hym.: Cynipidae). *Butlletí de la Institució Catalana d'Història Natural*, **78**, 3–9.

Pujade-Villar, J., García-Martiñón, R.D., Equihua-Martínez, A. & Estrada-Venegas, E. (2016) *Neuroterus fusifex* Pujade-Villar and Ferrey-Suay n. sp.(Hymenoptera: Cynipidae): first record of galls on catkins in Mexico. *Folia Entomológica Mexicana*, **2**, 75–83.

Pujade-Villar, J., Tovar, D.C., Ruíz, U.M.B., Martínez, A.E., Venegas, E.G.E. & Durán-Román, A.G. (2017) Una nueva especie de Neuroterus Hartig, 1840 de México (Hym., Cynipidae). *Butlletí de la Institució Catalana d'Història Natural*, **81**, 67–74.

Pujade-Villar, J., Cibrián-Tovar, D., Barrera-Ruíz, U.M. & Cuesta-Porta, V. (2018) Las especies de Neuroterus (Hartig, 1840) de México, con la descripción de una especie nueva (Hym., Cynipidae). *Entomología mexicana*, **5**, 453–461.

R Core Team (2020) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. Available at: https://www.R-project.org/. [Accessed 5th September 2021].

Rhie, A., McCarthy, S.A. & Fedrigo, O. *et al.* (2021) Towards complete and error-free genome assemblies of all vertebrate species. *Nature*, **592**, 737–746. https://doi.org/10.1038/s41586-021-03451-0

Rokas, A., Atkinson, R.J., Nieves-Aldrey, J.L., West, S.A. & Stone, G.N. (2002) The incidence and diversity of Wolbachia in gallwasps (Hymenoptera; Cynipidae) on oak. *Molecular Ecology*, **11**, 1815–1829.

Rokas, A., Melika, G., Abe, Y., Nieves-Aldrey, J.-L., Cook, J.M. & Stone, G.N. (2003) Lifecycle closure, lineage sorting, and hybridization revealed in a phylogenetic analysis of European oak gallwasps (Hymenoptera: Cynipidae: Cynipini) using mitochondrial sequence data. *Molecular Phylogenetics and Evolution*, **26**, 36–45.

Ronquist, F., Nieves-Aldrey, J.L., Buffington, M.L., Liu, Z., Liljeblad, J. & Nylander, J.A. (2015) Phylogeny, evolution and classification of gall wasps: the plot thickens. *PLoS One*, **10**, e0123301.

Rosenfeld, J.A., Foox, J. & DeSalle, R. (2016) Insect genome content phylogeny and functional annotation of core insect genomes. *Molecular Phylogenetics and Evolution*, **97**, 224–232.

Rozen, S., Skaletsky, H., Marszalek, J.D., Minx, P.J., Cordum, H.S., Waterston, R.H. *et al.* (2003) Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature*, **423**, 873–876.

Rundell, R.J. & Leander, B.S. (2010) Masters of miniaturization: convergent evolution among interstitial eukaryotes. *BioEssays*, **32**, 430–437.

Samacá-Sáenz, E., Egan, S.P. & Zaldívar-Riverón, A. (2020) Species diversity in the braconid wasp genus Allorhogas (Doryctinae) associated with Cynipid galls on live oaks (Quercus: Fagaceae) using natural history, Phylogenetics, and morphology. *Insect Systematics and Diversity*, **4**, 3.

Sanver, D. & Hawkins, B.A. (2000) Galls as habitats: the inquiline communities of insect galls. *Basic and Applied Ecology*, **1**, 3–11.

Scheffers, B.R., Joppa, L.N., Pimm, S.L. & Laurance, W.F. (2012) What we know and don't know about Earth's missing biodiversity. *Trends in Ecology & Evolution*, **27**, 501–510.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. & Zdobnov, E.M. (2015) BUSCO: assessing genome assembly and

annotation completeness with single-copy orthologs. *Bioinformatics*, **31**, 3210–3212.

Slater, G.S.C. & Birney, E. (2005) Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, **6**, 1–11.

Smit, A.F.A, Hubley, R. & Green, P. (2015) RepeatMasker Open-4.0. 2013–2015. Available at: http://www.repeatmasker.org [Accessed on 24th September 2020].

Smith, M.A., Rodriguez, J.J., Whitfield, J.B., Deans, A.R., Janzen, D.H., Hallwachs, W. & Hebert, P.D. (2008) Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. *Proceedings of the National Academy of Sciences*, **105**, 12359–12364.

Sork, V.L. (1993) Evolutionary ecology of mast-seeding in temperate and tropical oaks (Quercus spp.). *Frugivory and Seed Dispersal: Ecological and Evolutionary Aspects*, pp. 133–147. Springer, Dordrecht.

Srivathsan, A., Hartop, E., Puniamoorthy, J., Lee, W.T., Kutty, S.N., Kurina, O. & Meier, R. (2019) Rapid, large-scale species discovery in hyperdiverse taxa using 1D MinION sequencing. *BMC Biology*, **17**, 1–20.

Staats, M., Erkens, R.H., van de Vossenberg, B., Wieringa, J.J., Kraaijeveld, K., Stielow, B. *et al.* (2013) Genomic treasure troves: complete genome sequencing of herbarium and insect museum specimens. *PLoS One*, **8**, e69189.

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S. & Morgenstern, B. (2006) AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Research*, **34**, W435–W439.

Stathos, A. & Fishman, L. (2014) Chromosomal rearrangements directly cause underdominant F1 pollen sterility in *Mimulus lewisii–Mimulus cardinalis* hybrids. *Evolution*, **68**, 3109–3119.

Stone, G.N., Atkinson, R.J., Rokas, A., Aldrey, J.L.N., Melika, G., Acs, Z. *et al.* (2008) Evidence for widespread cryptic sexual generations in apparently purely asexual Andricus gallwasps. *Molecular Ecology*, **17**, 652–665.

Stork, N.E. (2018) How many species of insects and other terrestrial arthropods are there on earth? *Annual Review of Entomology*, **63**, 31–45.

Tedesco, P.A., Bigorne, R., Bogan, A.E., Giam, X., Jézéquel, C. & Hugueny, B. (2014) Estimating how many undescribed species have gone extinct. *Conservation Biology*, **28**, 1360–1370.

USDA. 2021. National Agricultural Statistics Service. Available at: https://data.nal.usda.gov/dataset/nass-quick-stats [Accessed on 20th April 2021].

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G. *et al.* (2001) The sequence of the human genome. *Science*, **291**, 1304–1351.

Wang, L.-G., Lam, T.T.-Y., Xu, S., Dai, Z., Zhou, L., Feng, T. *et al.* (2020) Treeio: an R package for phylogenetic tree input and output with richly annotated and associated data. *Molecular Biology and Evolution*, **37**, 599–603.

Ward, A.K., Sheikh, S.I. & Forbes, A.A. (2020) Diversity, host ranges, and potential drivers of speciation among the inquiline enemies of oak gall wasps (Hymenoptera: Cynipidae). *Insect Systematics and Diversity*, **4**, 3.

Weinersmith, K.L., Forbes, A.A., Ward, A.K., Brandão-Dias, P.F.P., Zhang, Y.M. & Egan, S.P. (2020) Arthropod community associated with the asexual generation of *Bassettia pallida* (Hymenoptera: Cynipidae). *Annals of the Entomological Society of America*, **113**, 373–388.

Weissensteiner, M.H., Bunikis, I., Catalán, A., Francoijs, K.J., Knief, U., Heim, W. *et al.* (2020) Discovery and population genomics of structural variation in a songbird genus. *Nature Communications*, **11**, 1–11.

Weld, L.H. (1921) American gallflies of the family Cynipidae producing subterranean galls on oak. *Proceedings of the United States National Museum*, **59**, 187–246.

Werren, J.H., Richards, S., Desjardins, C.A., Niehuis, O., Gadau, J., Colbourne, J.K. & Nasonia Genome Working Group (2010) Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science*, **327**, 343–348.

Wilson, A.D., Lester, D.G. & Edmonson, R.E. (2000) Live oaks, new hosts for *Odontocynips nebulosa* Kieffer (Hymenoptera: Cynipidae) in North America. *Proceedings of the Entomological Society of Washington*, **102**, 360–373.

Xu, X., Liu, F., Chen, J., Li, D. & Kuntner, M. (2015) Integrative taxonomy of the primitively segmented spider genus Ganthela (Araneae: Mesothelae: Liphistiidae): DNA barcoding gap agrees with morphology. *Zoological Journal of the Linnean Society*, **175**, 288–306.

Yahav, T. & Privman, E. (2019) A comparative analysis of methods for de novo assembly of hymenopteran genomes using either haploid or diploid samples. *Scientific Reports*, **9**, 1–10.

Yamada, T.K., Kitamura, S., Abe, S., Tajima, Y., Matsuda, A., Mead, J.G. & Matsuishi, T.F. (2019) Description of a new species of beaked whale (Berardius) found in the North Pacific. *Scientific Reports*, **9**, 1–14.

Yu, G., Smith, D.K., Zhu, H., Guan, Y. & Lam, T.T.Y. (2017) Ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, **8**, 28–36.

Zhang, Y.M., Gates, M.W. & Shorthouse, J.D. (2014) Testing species limits of Eurytomidae (Hymenoptera) associated with galls induced by Diplolepis (Hymenoptera: Cynipidae) in Canada using an integrative approach. *The Canadian Entomologist*, **146**, 321–334.

Zhang, L., Hood, G.R., Ott, J.R. & Egan, S.P. (2019) Temporal isolation between sympatric host plants cascades across multiple trophic levels of host-associated insects. *Biology Letters*, **15**, 20190572.

Zhang, Y.M., Egan, S.P., Driscoe, A.L. & Ott, J.R. (2021) One hundred and sixty years of taxonomic confusion resolved: *Belonocnema* (Hymenoptera: Cynipidae: Cynipini) gall wasps associated with live oaks in the USA. *Zoological Journal of the Linnean Society*, **191**, zlab001.

Zhu, D.H., Liu, Z., Lu, P.F., Yang, X.H., Su, C.Y. & Liu, P. (2015) New gall wasp species attacking chestnut trees: Dryocosmus zhuili n. sp.(Hymenoptera: Cynipidae) on Castanea henryi from southeastern China. *Journal of Insect Science*, **15**, 156.

Accepted 7 September 2021