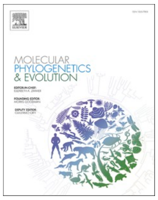
Molecular Phylogenetics and Evolution 187 (2023) 107807

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev



First phylogenetic analysis of the Nearctic madicolous midges of the genus *Androprosopa* Mik (Diptera: Thaumaleidae)

Robert J. Pivar[a], John K. Moulton[a,*], Bradley J. Sinclair[b]

[a] The University of Tennessee, Department of Entomology and Plant Pathology, 2505 E.J. Chapman Drive, 370 Plant Biotechnology Building, Knoxville, TN 37996, USA
[b] Canadian National Collection of Insects and Canadian Food Inspection Agency, K.W. Neatby Building, C.E.F., 960 Carling Avenue, Ottawa, Ontario K1A 0C6, Canada

ARTICLE INFO

Keywords:
Diptera
Thaumaleidae
Nearctic
Androprosopa
Phylogeny
Molecular

ABSTRACT

Molecular phylogenetic analyses were conducted to infer relationships between the eastern and western Nearctic *Androprosopa* Mik and amongst the considerably more diverse western Nearctic species. Fresh, molecular-grade material was obtained for all Nearctic *Androprosopa* species except two Mexican species, *An. sonorensis* (Arnaud & Boussy) and *An. zempoala* Sinclair & Huerta, that eluded capture. Molecular sequences from two nuclear protein-coding genes, big zinc finger (BZF) and molybdenum cofactor sulfurase (MCS), were sampled from representatives of several outgroup and ingroup taxa and analyzed phylogenetically using maximum likelihood criteria to confirm identifications of females and immatures using a barcoding approach, test species boundaries among morphologically similar species, and infer relationships among more morphologically disparate groups. Resulting phylogenies suggest the following with significant node (bootstrap) support: (1) the eastern Nearctic *Androprosopa* species form the sister group to the lineage comprised of all sampled Palearctic thaumaleids, i.e., *An. larvata* (Mik), *An. striata* (Okada), and *Thaumalea testacea* Ruthe; (2) the aforementioned lineage is the sister group to the clade comprised of western Nearctic *Androprosopa* species; (3) the western Nearctic *Androprosopa* species form three multispecies lineages, two of which can be further divided into three or more well founded species groups. Our results suggest that *Androprosopa* as currently defined is paraphyletic. Additionally, we propose several new species groups within the western Nearctic *Androprosopa* based on molecular and morphological data.

1. Introduction

Thaumaleidae, or madicolous midges, are a family of aquatic Diptera. They are small (2–5 mm), stocky flies that are typically brown or black, though some may be dark orange or shiny yellow. They are found worldwide on all continents except Antarctica (Vaillant 1977; Sinclair & Saigusa 2002). Thaumaleidae includes 202 described species, currently classified within seven genera (Sinclair et al. 2022). The immature stages are morphologically specialized and restricted to madicolous habitats (Vaillant 1956; Sinclair & Marshall 1987; Sinclair & Stuckenberg 1995). These habitats are characterized by thin films (2 mm or less) of flowing water over various substrates (e.g., rock, mud, moss or leaves) (Vaillant 1956; Mackie 2004; Pivar et al. 2021). Examples of such habitats include rock-face seepages, cascading streams, splash zones near waterfalls, and boulders in fast moving streams (Sinclair 2000).

Due to their restricted habitat and weak dispersal ability of the adults, thaumaleids are a rarely encountered and thus understudied group. The best and most efficient method for collecting adult Thaumaleidae is to sweep riparian vegetation near the madicolous substrate. If sweeping is unsuccessful, inspect the madicolous substrate for adults resting on moss, in cracks and crevices, or on the rock face itself, where they may be aspirated. Larvae may be removed from the substrate using forceps or washed into a white pan (Sinclair & Saigusa 2002).

The Nearctic Region includes three genera (*Androprosopa* Mik, *Thaumalea* Ruthe and *Trichothaumalea* Edwards) and 32 described species (Pivar et al. 2018). *Androprosopa* is the most speciose genus in the Nearctic Region, with 28 species. The earliest taxonomic work on Nearctic Thaumaleidae dates to Bezzi (1913), in which he described the first Nearctic species, *Androprosopa americana* (Bezzi) from Ithaca, New York. The two remaining eastern Nearctic species are *An. thornburghae* (Vaillant) and *An. vaillantiana* Sinclair, which were revised by Sinclair (1996) including keys and descriptions of life stages.

Androprosopa is far more diverse in the western Nearctic Region with

---

* Corresponding author.
E-mail addresses: rpivar@vols.utk.edu (R.J. Pivar), jmoulton@utk.edu (J.K. Moulton), bradley.sinclair@inspection.gc.ca (B.J. Sinclair).

https://doi.org/10.1016/j.ympev.2023.107807
Received 15 February 2023; Received in revised form 7 April 2023; Accepted 29 April 2023

25 described species (Pivar et al. 2018). Arnaud & Boussy (1994) described 17 new species, while Pivar et al. (2018) added three additional species and provided updated distribution maps, an identification key to males, and proposed six new species groups based on male genitalic characters. Females, larvae and pupae have been described for only a small number of species, but they still pose a problem for identification as they lack distinguishing characters (Arnaud & Boussy 1994; Gillespie et al. 1994). Species determination of females and immatures by association is problematic, as often times habitats support multiple species (Pivar et al. 2018).

Relationships amongst the Nearctic species of *Androprosopa* remain uncertain. The eastern species are characterized by short, blunt apical gonostylar spines, tapered gonostyli, and dense setae on the inner face of the gonocoxites, features shared with *An. ericfisheri* (Arnaud & Boussy) (Sinclair 1996). All western Nearctic species groups have tapered gonostyli terminating in apical or subapical claw-like setae, except for the *An. anolo* group (Pivar et al., 2018). Despite all of the studies on Nearctic Thaumaleidae, they remain an under collected and understudied group with much still to be learned about diversity and relationships at all levels of classification.

In Europe, the division of Thaumaleidae into species groups based on male terminalia was first attempted by Edwards (1929), who recognized two European genera and three broad groupings. Species groups and generic definitions were reinvestigated by Vaillant (1978), who recognized four European genera and proposed five species groups for the genus *Thaumalea*. More recently, Wagner (2002) recognized three European genera and divided the genus *Thaumalea* into 13 species groups, along with several provisional groups. Haubrock et al. (2017) included species from seven of these 13 species groups, and based on this limited dataset, two species groups were found to be non-monophyletic. Much work remains to investigate the monophyly of the European fauna and their relationship to other segregates within the family.

Over the past 170 + years, knowledge about thaumaleids has grown primarily through descriptive work and little has been done with regard to exploring evolutionary relationships. The only phylogenetic analysis within the family is that of Haubrock et al. (2017), which examined divergence and dispersal of 14 European species. Arnaud & Boussy (1994) made reference to species they hypothesized may be closely related to one another in the species diagnoses, but they did not investigate relationships formally. Sinclair (1996) proposed hypotheses about the affinities between certain Nearctic taxa, but again, scope was narrow, encompassing only a few species. Pivar et al. (2018) proposed species groups on the basis of male genitalic characters. The purpose of this paper is to report results from the first comprehensive phylogenetic analysis of Nearctic species of *Androprosopa*. Widespread species, like *An. gillespieae* (Arnaud & Boussy) and *An. americana*, occur in disparate habitat types across their ranges, which often indicates species complexes (Moulton & Adler 1995). By sampling multiple individuals of these and other morphospecies across their known ranges, we were able to test species limits and infer inter-species relationships, including assessment of whether the eastern Nearctic fauna is more closely related to the Palearctic or western Nearctic faunas. Both of the genes utilized here are also well suited to function as barcode fragments useful in associating unknown or currently inseparable life stage of any Nearctic species to almost exclusively name-bearing males.

2. Materials and methods

2.1. Taxon sampling

Field collection over the past decade has yielded molecular grade material for all but two species, *An. sonorensis* (Arnaud & Boussy) and *An. zempoala* Sinclair & Huerta, of described Nearctic *Androprosopa*. Upon collection, specimens were placed immediately into 95% ethanol and later stored frozen at −20 °C until use. At least one exemplar for each available species was used for molecular study (Table 1). Wherever possible or appropriate, specimens of each morphospecies were sequenced from disparate populations to capture intraspecific variation and potentially detect cryptic species. When possible, the type locality of each species was collected to obtain topotypical material.

Distal outgroups (Table 2) included members of Culicidae and representatives from the Simuliidae, the sister group to Thaumaleidae (Moulton 2000; Bertone et al. 2008; Borkent 2012). Proximal outgroups included *Afrothaumalea stuckenbergi* Sinclair, *Trichothaumalea elakalensis* Sinclair and *Thaumalea testacea* Ruthe. Additionally, two Palearctic species of *Androprosopa*, *An. larvata* Mik and *An. striata* (Okada), were included.

2.2. Genomic sampling

Gene regions from two nuclear-coding loci — big zinc finger (BZF) and molybdenum cofactor sulfurase (MCS) — were chosen to infer phylogenetic relationships. These loci were chosen due to their presence as single copies, large size, high levels of nonsynonymous divergence, and demonstrated ability to reconstruct relationships within comparable taxonomic groups (Senatore et al. 2014; Winkler et al. 2015). In lieu of an obviously more powerful but prohibitively costly phylogenomic approach employing dozens or more loci, this study relies upon weighing topological concordance and nonparametric bootstrap node support across natural partitions of these two powerful markers.

2.2.1. Big zinc finger (BZF)

The potential of BZF as a phylogenetic marker was first demonstrated by Senatore et al. (2014), who utilized it to reconstruct a molecular phylogeny within the *Simulium jenningsi* species group (Moulton & Adler 1995). Given its success within the closely related family Simuliidae, BZF seemed an ideal choice for use in this project. Dipterans are one of several eukaryotic organisms that have single copies of BZF, a rapidly evolving DNA-binding transcriptional regulator (Senatore et al. 2014). Although BZF exceeds three kilobases of coding region, primers used herein focused on the distal two-thirds of the gene, resulting in a roughly 2-kilobase region containing only a single intron in thaumaleids.

2.2.2. Molybdenum cofactor sulfurase (MCS)

MCS was used for its high degree of phylogenetic informativeness when compared against several other nuclear genes, as illustrated by Winkler et al. (2015) and Blaschke et al. (2018). Primers were designed to amplify and sequence a roughly 2.5-kb pair region of the gene. This region, with 7 introns present in thaumaleids, was more intron burdened than optimal, leading to direct sequencing difficulties resulting from intron length polymorphisms in several taxa. This was remedied using complimentary internal sequencing primers designed from adjacent exonic regions or simply truncation if difficult introns were located near the 5′ or 3′ end of the amplicon.

2.3. Laboratory methods

Specimens of each taxon (either whole or terminalia only) were used to extract total DNA using either Thermo Scientific's GeneJET Genomic DNA Purification Kit #K0722 (ThermoScientific, Waltham, MA) or Omega Bio-tek's E.Z.N.A.® Insect DNA Kit #D0926-02 (Omega Bio-tek, Inc., Norcross, Georgia) following the manufacturers' suggested protocol except for reducing the final elution volume to a total of 100 μL. The resulting gDNA was stored at −20 °C and the cleared voucher specimen stored in 70% ethanol. Amplifications were performed in GenePro (Bioer Technology Co., Hangzhou, China) thermal cyclers, using 50 μL PCR reactions filled with TaKaRa Ex Taq Hotstart DNA polymerase (Takara Bio, Shiga, Japan) kit components per the manufacturer's suggested protocol, with 1.5 μL of template DNA, and 3 μL of 20 μM custom forward and reverse primers (Table 3). Less to completely nondegenerate taxon- or clade-specific internal primers were sometimes designed and used to obtain sufficient yields of DNA via reamplification. These

Table 1
Ingroup exemplars of Nearctic *Androprosopa* Mik studied with species name, geographical source data, genomic sampling, GenBank accession numbers, and fragment size in analyzed data matrix.

Species	Specimen Locality Information	Locality Coordinates	Genbank Accession # and # of base pairs (bp)	
			BZF	**MCS**
***Androprosopa americana* (Bezzi)**	USA: GA: Towns Co.: Brasstown Bald, seep near waterfall	34°51′36″ −83°48′09″	MK190468 2848 bp	MK206974 1255 bp
	USA: MS: Tishomingo Co.: Tishomingo State Park, Outcroppings Trail, waterfall	34°36′32″ −88°10′31″	N/A	MK206973 2020 bp
	CAN: ON: Owen Sound, Inglis Falls Conservation Area, springs/seeps	44°31′37″ −80°56′07″	N/A	MK206975 2111 bp
***An. anolo* (Schmid)**	USA: OR: Marion Co.: Willamette N.F., French Crk. Rd, trib.	44°46′58″ −122°11′12″	MK190490 851 bp	MK206992 2049 bp
	USA: WA: Mt. Rainier N.P., E. of Christine Falls, roadside seep	46°46′47″ −121°46′21″	MK190489 1865 bp	N/A
***An. apache* Pivar & Moulton**	USA: AZ: Apache Co.: Apache-Sitgreaves N.F., Greer, Squirrel Spring	34°02′27″ −109°27′32″	MK190503 2624 bp	MK207002 2031 bp
***An. arnaudi* Pivar**	USA: AZ: Graham Co.: Graham Mtns., Wet Canyon	32°39′06″ −109°48′48″	MK190502 2652 bp	MK207001 1916 bp
***An. becca* (Arnaud & Boussy)**	CAN: BC: East Kootenay, Hwy. 3/95, roadcut seep	49°10′18″ −116°00′31″	MK190481 2068 bp	MK206986 2052 bp
	USA: MT: Lincoln Co.: Hwy. 37, roadcut seeps	48°41′24″ −115°18′49″	MK190482 886 bp	N/A
	USA: WA: Mt. Rainier N.P., small falls E of Christine Falls next to road	46°46′51.2″ −121°46′45″	N/A	MK206985 2048 bp
***An. brothersi* (Arnaud & Boussy)**	USA: CA: Santa Clara Co.: Uvas Canyon County Park, ex. sweeps along Swanson Crk.	37°05′03″ −121°47′34″	MK190475 2167 bp	MK206980 2077 bp
	USA: CO: Mineral Co.: San Juan N.F., Hwy. 160, Falls Crk. nr. parking lot	37°26′31″ −106°52′37″	MK190477 484 bp	MK206982 967 bp
	USA: SD: Pennington Co.: Black Hills N.F., Eleven Hour Gulch	44°23′38″ −103°54′42″	MK190476 1970 bp	MK206981 1984 bp
***An. buckae* (Arnaud & Boussy)**	USA: CA: Humboldt Co.: Trinidad State Beach, sea cliff face	41°03′25″ −124°09′00″	MK190497 2048 bp	MK206997 1962 bp
	USA: CA: Sierra Co.: Big Spring Falls, seepage area	39°35′48″ −120°36′38″	MK190495 1929 bp	N/A
	USA: CA: Siskiyou Co.: Hedge Crk., splash zones in crk.	41°14′14″ −122°16′09″	MK190496 908 bp	MK206996 2026 bp
***An. chandlerorum* (Arnaud & Boussy)**	USA: CA: Nevada Co.: Carpenter Ridge, Cirque Lake and stream	39°25′01″ −120°18′33″	MK190491 2675 bp	MK206993 1955 bp
***An. coloradensis* (Arnaud & Boussy)**	USA: CO: San Juan Co.: San Juan N.F., Hwy. 550, ex. Coal Crk. waterfalls on W side of road	37°42′29″ −107°46′17″	MK190478 2744 bp	MK206983 1977 bp
***An. confracta* (Arnaud & Boussy)**	USA: OR: Linn Co.: Santiam S.F., Monument Peak Rd., 820 m	44°42′34″ −122°22′16″	MK190473 1928 bp	MK206979 1948 bp
	USA: WA: Mt. Baker-Snoqualmie N.F., Marblemount, Cascade R. Rd., roadside seep	48°32′10″ −121°17′47″	MK190474 784 bp	N/A
***An. elnora* (Dyar & Shannon)**	USA: WA: Spokane Co.: Mt. Spokane SP Deadman Crk., 1325 m	47°54′20″ −117°06′35″	MK190470 2715 bp	MK206977 2055 bp
	USA: WA: Mt. Rainier N.P., below Comet Falls, streams/seeps	46°47′29″ −121°46′55″	MK190471 897 bp	N/A
***An. ericfisheri* (Arnaud & Boussy)**	CAN: BC: Golden, Quartz Crk. Rd., trickling stream	51°27′55″ −117°21′03″	MK190469 2786 bp	MK206976 1877 bp
***An. falciformis* (Arnaud & Boussy)**	USA: ID: Latah Co.: Moscow Mt., 1000 m, ex. cascading str.	46°47′43″ −116°54′05″	MK190484 2055 bp	MK206988 2018 bp
***An. fusca* (Garrett)**	CAN: BC: Columbia-Shuswap, crk.	51°12′29″ −118°12′09″	MK190488 2017 bp	MK206991 1882 bp
***An. gillespieae* (Arnaud & Boussy)**	CAN: BC: Hwy 23, 13 km N Revelstoke, 600 m	51°04′03″ −118°10′31″	MK190508 2772 bp	MK207007 1969 bp
	USA: CO: Lake Co.: Black Cloud Crk., splash zones	39°04′10″ −106°26′12″	MK190506 663 bp	MK207005 2044 bp
	USA: UT: Utah Co.: Mt. Timpanogos alpine loop Rd., above Sundance, ex. cascading stream next to road	40°23′41″ −111°34′56″	MK190507 2641 bp	MK207006 1962 bp
	USA: WA: Mt. Rainier N.P., Sunrise Bridge	46°54′27″ −121°32′32″	MK190505 872 bp	MK207004 1958 bp
***An. idahoensis* (Arnaud & Boussy)**	USA: ID: Nez Perce Co.: Juliaetta Falls	46°31′22″ −116°44′49″	MK190493 1957 bp	MK206995 1950 bp
	USA: CA: Tahoe N.F., waterfall 4 miles E of Downieville	39°33′31″ −120°45′40″	MK190494 839 bp	N/A
***An. lindsayorum* (Arnaud & Boussy)**	USA: CA: Del Norte Co.: Redwood N.P., Enderts beach, sea cliff seep	41°42′21″ −124°08′33″	MK190480 892 bp	N/A
	USA: WA: Mt. Rainier N.P., E of Nisqually R. bridge, roadcut seeps	46°46′48″ −121°45′43″	MK190479 2580 bp	MK206984 2013 bp
***An. melanderi* (Arnaud & Boussy)**	USA: CA: Nevada Co.: Culvert along Sagehen Crk.	39°26′04″ −120°16′52″	MK190472 2043 bp	MK206978 1792 bp
***An. palouse* (Arnaud & Boussy)**	CAN: BC: Cranbrook, Gold Creek Rd, ex. small crk., splash zone	49°27′06″ −115°41′15″	MK190500 2022 bp	MK206999 1860 bp

(*continued on next page*)

Table 1 (*continued*)

Species	Specimen Locality Information	Locality Coordinates	Genbank Accession # and # of base pairs (bp)	
			BZF	MCS
	USA: CO: Clear Creek Co.: Cascade Crk.	39°42′28″ −105°36′35″	MK190499 874 bp	N/A
	USA: ID: Nez Perce Co.: Lake Waha & Craig Mtn., stream	46°11′41″ −116°49′12″	MK190501 866 bp	MK207000 1980 bp
	USA: WY: Bridger-Teton N.F., creek	42°32′46″ −110°44′18″	MK190498 1989 bp	MK206998 2015 bp
An. rainierensis **Pivar & Sinclair**	USA: WA: Pierce Co.: Mt. Rainier N.P., Narada Falls to Reflection Lake tr., small stream	46°46′15″ −121°44′44″	MK190483 1961 bp	MK206987 2031 bp
An. santaclaraensis **(Brothers)**	USA: CA: Santa Clara Co.: Uvas Canyon County Park, sweeps along Swanson Crk.	37°05′03″ −121°47′34″	MK190492 2013 bp	MK206994 2049 bp
An. schmidiana **(Arnaud & Boussy)**	USA: CA: Nevada Co.: Big Culvert along Sagehen Crk., Malaise trap	39°26′04″ −120°16′52″	MK190510 2739 bp	MK207009 1963 bp
	USA: CA: Nevada Co.: Tahoe N.F., River Rd. Hwy. 89, crk. above waterfall	39°15′56″ −120°12′37″	MK190511 854 bp	MK207010 1714 bp
	USA: CA: Santa Clara Co.: Uvas Canyon County Park, sweeps along Swanson Crk.	37°05′03″ −121°47′34″	MK190509 2612 bp	MK207008 1985 bp
	USA: CA: Sierra Co.: Hwy. 49, crk.	39°34′00″ −120°39′50″	MK190512 1930 bp	MK207011 1999 bp
	USA: CA: Siskiyou Co.: Beaver Crk., splash zones	41°56′13″ −122°49′18″	MK190513 1954 bp	MK207012 964 bp
	USA: NV: White Pine Co.: Great Basin N.P., Baker Crk., ex. splash zones at trailhead	38°58′36″ 114°14′44″	MK190504 2655 bp	MK207003 1896 bp
An. thornburghae **(Vaillant)**	USA: GA: Towns Co.: Brasstown Bald, seep near waterfall	34°51′36″ −83°48′09″	MK190466 2780 bp	MK206971 1836 bp
An. vaillantiana **Sinclair**	USA: NC: Swain Co.: Great Smoky Mountains N.P., Clingmans Dome, seep at parking lot	35°33′24″ −83°29′39″	MK190467 2352 bp	MK206972 1959 bp
An. waha **(Arnaud & Boussy)**	USA: WA: Spokane Co.: Mt. Spokane S.P., Deadman Crk., 1325 m	47°54′20″ −117°06′35″	MK190486 2707 bp	MK206990 1955 bp
	USA: CA: San Bernardino Co.: Forest Falls, along Falls Crk., 1873 m	34°05′07″ −116°53′44″	MK190487 906 bp	N/A
An. zelmae **(Arnaud & Boussy)**	USA: WA: Mt. Rainier N.P., Comet Falls Tr., before falls overlook, 1456 m, seep	46°47′39″ −121°46′45″	MK190485 2005 bp	MK206989 1950 bp

Table 2
Exemplars of outgroup taxa studied with species name, geographical source data, genomic sampling, GenBank accession numbers and fragment size in analyzed data matrix.

Family	Species	Specimen Locality Information	Genbank Accession # and # of base pairs (bp)	
			BZF	MCS
Culicidae	*Aedes aegypti* (Linnaeus)	N/A	XM_001654141.2 1195 bp	N/A
	Anopheles gambiae Giles	N/A	XM_319733.4 1269 bp	XM_310528.5 1857 bp
	Culex quinquefasciatus Say	N/A	XM_001861745.1 1239 bp	XM_001870703.1 1905 bp
Simuliidae	*Cnephia ornithophilia* Davies, Peterson & Wood	USA: VA: Blackwater River	N/A	MK207013 2010 bp
	Greniera humeralis Currie, Adler & Wood	USA: CA: Lake Co.	N/A	MK207014 1958 bp
	Parahelodon gibsoni (Twinn)	CAN: ON: Algonquin Park	N/A	MK207015 900 bp
	Parasimulium stonei Peterson	USA: OR: Aufderheide Rt. 19	MK190514 2281 bp	N/A
	Parasimulium crosskeyi Peterson	USA: OR: Multnomah Falls	MK190515 2463 bp	N/A
	Simulium donovani Vargas	USA: AZ: Pima Co.	N/A	MK207017 909 bp
	Simulium krebsorum Moulton & Adler	USA: SC: Harman Mill Pond	JX847702 2645 bp	MK207016 938 bp
Thaumaleidae	*Afrothaumalea stuckenbergi* Sinclair	RSA: Western Cape: Cederberg Wilderness, Duiwelsgat, 693 m	MK190516 2713 bp	MK207018 1981 bp
	Androprosopa larvata (Mik)	DEU: Hesse: Rhon Biosphere Reserve, N. slope of Heidelstein	MK190518 2545 bp	MK207020 2031 bp
	Androprosopa striata (Okada)	JPN: Kyushu, Kusatsu godoni	MK190519 2715 bp	MK207021 2022 bp
	Thaumalea testacea Ruthe	FRA: Vosges, Bussang	MK190520 2714 bp	MK207022 2033 bp
	Trichothaumalea elakalensis Sinclair	USA: WV: Blackwater Falls S.P., Elakala Falls	MK190517 2902 bp	MK207019 2000 bp

Table 3
Primers[a] used in this study.

Gene	Primer	F/R	Sequence (5′ →3′)[b]	Length
BZF	90F	F	GARTGYGARGTNACNGARGARGT	23-mer
Piece 1	94F (111F)[c]	F	GARGARGTNATHACNGAYGAYTGG	24-mer
	1.5F	F	CCNACNGAYCARGAYGARTA	20-mer
	Thaum 99F	F	GGNGARGARTGYGTNGARGT	20-mer
	Thaum P1R	R	TGYGATCGWTGTGGYATGKCNTT	23-mer
Piece 2	385F (330F) [c]	F	CCNTTYGTNTGYCARCARTGYGG	23-mer
	Thaum P2F	F	TGYGGTGARATGTTYTGGGAYCG	23-mer
	Int2F	F	ACNTGYGGNGAWATGTTYTGGGA	23-mer
	Int2F Rpl	F	GATTCNCCNTTTGARTGYGA	20-mer
	765R	R	TGYTGGAARGCNTTNGCNGAYGG	23-mer
	777R	R	AARCARGARCCNATHCAYACNGG	23-mer
	787R	R	GARAARCCRTAYGTRTGYTCSATNTG	26-mer
	787R Rpl1	R	TGYCCWCGKGCDTTCAAYCAAMG	23-mer
	Thaum787R Rpl2	R	CAYCAYTCRGCRCMRGAYACNGT	23-mer
Piece 3	Thaum P3F	F	TAYGGYTGYCGYTTYTGYTGG	21-mer
	New Thaum P3F	F	CARATHCAYATHCGYATYCAYAC	23-mer
	Thaum 1203R (1101R) [c]	R	TTNACNGGNGARCGNAGDATYTCYTC	26-mer
	New Thaum P3R	R	GAYAARTAYATGGAYAARTAYAAYAG	26-mer
MCS	55F(Y) [d]	F	TGYTAYYTNGAYCAYGCYGG	20-mer
Piece 1	55F(R) [d]	F	TGYTAYYTNGAYCAYGCRGG	20-mer
	78F	F	TNWWYTGYAAYCCNCAYAC	19-mer
	Thaum P1R	R	CAYATTTGTGGYGATTCMAAYGA	23-mer
Piece 2	446F [e]	F	TTYGTNGGNTTYGCYGARGT	20-mer
	463F [e]	F	YCTNCGNACNGGNTGYTTYTGYAA	24-mer
	464F	F	ACRGGRTGYTTYTGYAAYCC	20-mer
	838R [e]	R	TGYCARATGATHTGYATHGAYCA	23-mer
	840R [d]	R	ATGATHTGYATHGAYCARRSNACNGG	26-mer

[a] Sequences for additional, less degenerate, clade-specific PCR and sequencing primers used in this study are available upon request.
[b] R = A/G; Y = C/T; S = C/G; W = A/T; K = G/T; M = A/C; H = A/C/T; D = A/G/T; V = A/C/G; N = A/T/G/C.
[c] Primer names published by Senatore et al. (2014).
[d] Primer names published by Winkler et al. (2015).
[e] Primer names published by Blaschke et al. (2018), although slightly redesigned to better suit Thaumaleidae.

primer sequences are available from the authors upon request. The following PCR regimen was used: Initial 90 s denaturing step at 94 °C, then 4 cycles of 30 s at 94 °C, 30 s at 56 °C and 75 s at 72 °C, followed by 4 cycles of 30 s at 94 °C, 25 s at 52 °C and 75 s at 72 °C, 9 cycles of 30 s at 94 °C, 20 s at 48 °C and 75 s at 72 °C and finally, 38 cycles of 30 s at 94 °C, 20 s at 45 °C and 75 s at 72 °C. Once all cycling is complete the program ends with a 5 min 72 °C soak and 15 °C hold. PCR products were electrophoresed in a 1% agarose gel at 115 V for 30 min. Silica column-based gel purified PCR products served as template for Sanger® sequencing in both directions in 20 μL reactions using 20-fold diluted (=0.4 μL) BigDye® v3.1 terminators (Applied Biosystems, Carlsbad, California) utilizing a 5X dilution buffer blend containing ABI 5X dilution buffer, BDX-64 (MCLAB Products, South San Francisco, CA), and BetterBuffer (GelCompany, San Francisco, CA) and a cycling regimen that included a three-temperature (47, 45, 43 °C) touchdown annealing protocol. Sequencing reactions were cleaned using Centrisep columns (Princeton Separations, Adelphia, NJ) and dried in a Centrivap Concentrator (LABCONCO, Kansas City, MO). Dried samples were sent to the University of Tennessee, Knoxville, Genomics Core for sequencing. Sequences from opposing strands were reconciled and verified for accuracy using Sequencher 4.7 (Gene Codes Corp., Ann Arbor, MI). The GT-AG rule was followed to determine exon–intron boundaries (Rogers & Wall 1980) ensuring the presence of a resultant continuous open reading frame. Introns were excised in Sequencher prior to alignment and phylogenetic analysis. All generated sequences are deposited in GenBank as genomic DNA, i.e., introns included (accession numbers appear below).

2.4. *DNA alignment and phylogenetic analysis*

Reconciled, unaligned, and intron-free nucleotide sequences were exported from Sequencher as Nexus-formatted (Maddison et al. 1997) files for subsequent alignment using Mesquite 2.72 (Maddison & Maddison 2011). Alignments were created for each gene and a concatenated data set. For each, nucleotide sequences were translated to amino acids that were subsequently aligned using Opal 2.1.0 (Wheeler & Kececioglu 2007) as a plug-in. Using the align DNA to AA feature, amino acid alignments were used as a guide to create correspondingly aligned DNA matrices. Before analysis, natural partitions, i.e., codon positions, within each of the three data sets were designated. Phylogenetic analyses consisted of maximum likelihood analyses as implemented in IQ-TREE2 (Nguyen et al. 2015) and were performed on entire and third position-free nucleotide alignments of each gene and the concatenated data set. Models were chosen using IQ-TREE ModelFinder (Kalyaanamoorthy et al. 2017) according to AICc values. We assessed node support using Ultrafast bootstrap (UBS) (Hoang et al. 2018). Considering the risk of overestimating branch supports with UBS, parameter hill-climbing nearest neighbor interchange (-bnni) was implemented to reduce overestimation (Minh et al. 2020).

3. Results

3.1. *Data set properties*

3.1.1. *Big zinc finger (BZF)*

The final aligned matrix included 59 terminal taxa, including six distal outgroups representing three culicid (*Aedes*, *Anopheles*, *Culex*) and two simuliid (*Parasimulium*, *Simulium*) genera and 53 thaumaleids representing all proximal outgroup and successfully acquired ingroup taxa, eleven of the latter multiply sampled. Sequences ranged in length from 484 to 2902 nucleotides (Tables 1 & 2), with short-length sequences, i.e., those less than 1000 bp, originating from ingroup specimens barcoded for identification purposes. Generated BZF sequences were deposited in GenBank under accession numbers MK190468–MK190515.

Table 4
Maximum likelihood nonparametric bootstrap node support observed for each clade appearing in Fig. 1 and conflicting topologies observed from inferences of all six data sets (concatenated BZF & MCS all codon positions; concatenated BZF & MCS first & second codon positions only; BZF all codon positions; BZF first & second codon positions only; MCS all codon positions; MCS first & second codon positions only) analyzed. N/A Taxon sampling in data set precluded recovery of specified clade. Node numbers and bootstrap support formatted similarly in style and color to corresponding ones in aforementioned figure.

Node	Conc **All**	Conc 1 + 2	BZF all	BZF 1 + 2	*MCS all*	MCS 1 + 2	Node	Conc **All**	Conc 1 + 2	BZF all	BZF 1 + 2	*MCS all*	MCS 1 + 2
1	75	100	100	100	N/A	N/A	33	100	98	100	96	N/A	N/A
2	—	60	50	65	N/A	N/A	34	100	84	100	83	99	88
3	**100**	**100**	**100**	**100**	**100**	**100**	35	97	40	N/A	N/A	N/A	N/A
4	**78**	**100**	**98**	**100**	**100**	**100**	**36**	**100**	**100**	**100**	**100**	**66**	**100**
5	100	100	100	100	N/A	N/A	37	94	—	—	—	100	—
6	93	55	N/A	N/A	N/A	N/A	38	100	—	79	—	100	100
7	41	64	N/A	N/A	N/A	N/A	39	100	89	100	45	99	90
8	54	63	N/A	N/A	N/A	N/A	40	98	100	72	—	N/A	N/A
9	100	100	N/A	N/A	100	100	41	100	99	95	—	100	99
10	**100**	**100**	**100**	**100**	**100**	**100**	42	93	N/A	93	—	N/A	N/A
11	**92**	**87**	*	*	**100**	**100**	**43**	**100**	**100**	**100**	**100**	**100**	**100**
12	**100**	**100**	**98**	**100**	**100**	**100**	**44**	**100**	**100**	**100**	**100**	**100**	**100**
14	99	95	91	92	95	66	**45**	**100**	**87**	**100**	**99**	**99**	**99**
15	100	100	100	100	100	100	46	100	—	100	—	N/A	N/A
16	100	100	100	97	100	100	47	**100**	**81**	**98**	—	**100**	**100**
17	100	100	100	100	100	100	48	94	—	—	42	N/A	N/A
18	100	100	N/A	N/A	100	97	**49**	**100**	**100**	**95**	****	**100**	**100**
19	87	95	N/A	N/A	85	81	50	100	100	96	****	100	100
20	98	68	84	78	97	100	51	100	79	100	61	N/A	N/A
21	**100**	**100**	**100**	**98**	**100**	**100**	52	91	—	70	—	68	—
22	100	100	100	100	N/A	N/A	53	100	—	100	—	100	—
23	100	98	94	82	98	95	54	100	—	100	—	100	—
24	100	100	100	75	100	100	55	100	100	100	46	100	97
25	**100**	**100**	**100**	**100**	**100**	**100**	56	41	—	—	—	—	48
26	**	**	72	**	56	**	57	100	—	68	—	100	98
27	100	100	100	100	NA	N/A	58	62	—	—	—	62	63
28	100	100	100	95	100	100	59	72	—	—	—	54	—
29	100	100	100	62	100	99	60	89	85	—	—	28	—
30	**100**	**100**	**100**	**99**	**100**	**99**	61	95	—	—	—	—	—
31	**56**	**90**	**81**	**58**	***	**75**	62	95	85	—	95	72	—
32	100	100	100	96	100	100	63	84	—	—	—	81	89

N/A Taxon sampling in data set precluded recovery of specified clade.
— Alternative poorly supported topologies observed.
*Alternate topology: *Afrothaumalea stuckenbergi* sister to *Trichothaumalea elakalensis* (98% and 94% bootstrap support in BZFall and BZFnt1 + 2 analyses, respectively).
**Alternate topology: *An. melanderi* sister to *An. brothersi* (ConAll & Connt1 + 2, 42% & 79% bootstrap support, respectively; MCSnt1 + 2, 95% bootstrap support.
**Alternate topology: *An. melanderi* sister to *An. confracta* and *An. brothersi* (BZFall, 67% bootstrap support).
***Alternate topology: *An. coloradensis* sister to blue clade (66% bootstrap support).
****Alternate topology: *An. santaclaraensis* sister to *An. palouse* BC (86% bootstrap support).

3.1.2. Molybdenum cofactor sulfurase (MCS)

The final aligned matrix included 54 terminal taxa, including seven distal outgroups representing three culicid (*Anopheles*, *Culex*) and four simuliid (*Cnephia*, *Greneria*, *Parahelodon*, *Simulium*) genera and 47 thaumaleids representing the three proximal outgroups, the two Palearctic *Androprosopa* species, and all successfully acquired Nearctic *Androprosopa* species, seven of the latter multiply sampled. Sequences ranged in length from 900 to 2111 nucleotides (Tables 1 & 2), with short-length sequences, i.e., those less than 1000 bp, originating from specimens within ingroup species barcoded for identification purposes. Generated MCS sequences were deposited in GenBank under accession numbers MK206973–MK207013.

3.1.3. Concatenated data set

The final aligned matrix included 66 terminal taxa, including 10 distal outgroups representing three culicid (*Aedes*, *Anopheles*, *Culex*) and five simuliid (*Cnephia*, *Greneria*, *Parahelodon*, *Parasimulium*, *Simulium*) genera and 56 thaumaleids representing all successfully acquired proximal outgroup and ingroup taxa. Thirteen *Androprosopa* species, mostly western Nearctic species, were multiply sampled. The concatenated alignment, with gene partitions designated, is available from the authors upon request.

3.2. Phylogenetic analysis

Phylogenetic analysis of entire individual genes and nonsynonymous sites yielded largely congruent topologies generally lacking strongly supported discordance, prompting us to undertake a total evidence approach, which resulted in trees possessing enhanced node support in most circumstances compared to those from individual gene trees . The maximum likelihood tree (ln(L) = -71361.962; Selected model and parameters = TPM2{1.8594, 3.98293} + FQ + I{0.130344} + G4{0.985918}; Length = 6537) inferred from all codon positions of concatenated BZF and MCS sequences, with *An. santaclaraensis* (Brothers) dropped into an unresolved polytomy with two other clades due to conflicting resolutions, is presented as the best current hypothesis of relationships (Fig. 1). Differences in topology and node support between this tree and ones recovered from other analyses are summarized in Table 4 and are discussed below. Bootstrap support (BS) values of ≥ 70% were considered significant support.

Rooting recovered phylogenies with Culicidae yields a strongly supported Simuliidae (78–100%). Limited taxon sampling and differences in genomic sampling within Simuliidae preclude discussion about inter-relationships other than for those of the concatenated data, where trees show *Parasimulium* as sister to the remaining Simuliidae, with significant support (NS = 93%) found only with all positions considered. Neither tree recovered from analyses of the concatenated data supports

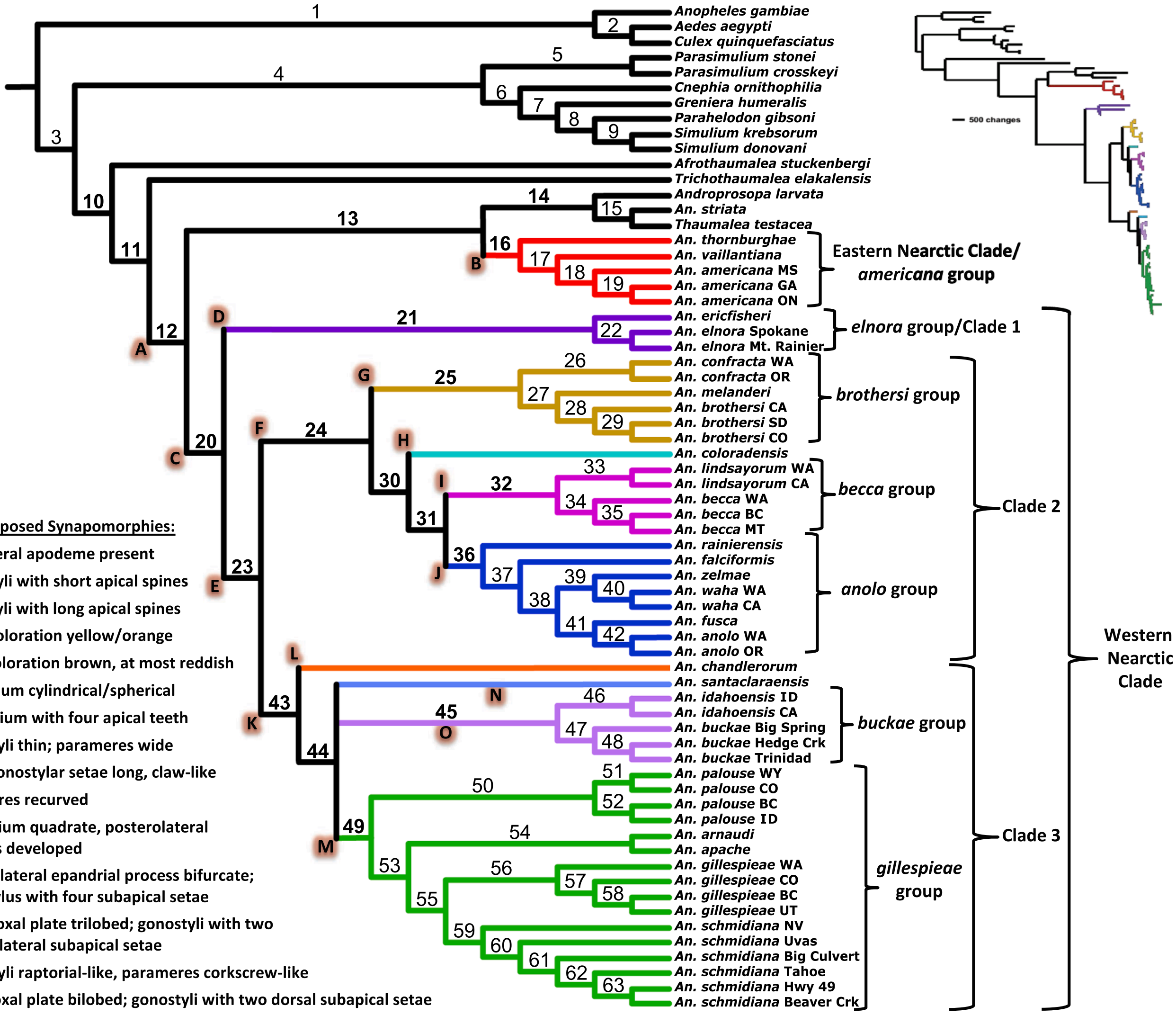


Fig. 1. Hypothesized phylogenetic relationships of Nearctic *Androprosopa* based on maximum likelihood analysis of all codon positions of the concatenated big zinc finger (BZF) and molydenum cofactor sulfurase (MCS) data sets with proposed morphological synapomorphies. *Androprosopa santaclaraensis* was dropped into polytomy to show discordance observed among inferences from various data sets. Black branches represent outgroup taxa while colored branches indicate ingroup taxa. State and provincial abbreviations or names of specific localities are provided for multiply sampled taxa. Bootstrap support observed for each node across all six analyses conducted is provided as Table 4. Inset (top right) is a phylogram of the maximum likelihood tree with taxon names removed and major clades similarly color-coded.

the current hypothesis of tribal splits within Simuliinae (Moulton, 2000; Adler et al., 2004).

Within Thaumaleidae, reconstructed trees differ most notably in the relationship between *Tr. elakalensis* and *Af. stuckenbergi* to one another and to the remaining Thaumaleidae, the relationship of *An. santaclaraensis* to the *An. buckae* and *An. gillespieae* species groups defined below, and finally, inter- and intraspecific relationships among exemplars within morphologically similar species groups, particularly those including *An. anolo* (Schmid), *An. buckae* (Arnaud & Boussy), *An. gillespieae*, and *An. melanderi* (Arnaud & Boussy). These discrepant topologies, their supporting data (by gene and partition) and node support are further discussed below.

Analyses of the concatenated data and MCS alone recovered *Af. stuckenbergi* as sister to *Tr. elakalensis* plus the remaining sampled Thaumaleidae with node support ranging from 87 to 100%. In analyses of BZF nucleotides, however, *Af. stuckenbergi* and *Tr. elakalensis* strongly (NS = 94 to 98%) pair and form the sister group to a maximally supported (NS = 100) clade comprising all remaining sampled Thaumaleidae.

All analyses recovered a clade comprised of *An. larvata* as sister to *An. striata* plus *Thaumalea testacea*, all of which in turn formed the sister group to the eastern Nearctic species in the following branching order: *An. thornburghae* + (*An. americana* + *An. vallantiana*). Node support for each node in this clade was highly robust (90+%) in most analyses, except ones of BZF alone where the node supporting the entire clade was 57% (all sites) and 69% (nonsynonymous sites). In all reconstructions this entire clade formed the strongly supported (NS = 90 to 100%) sister group to the Western Nearctic *Androprosopa* species.

Monophyly of the western Nearctic *Androprosopa* clade ranged from 48 to 100% bootstrap support, the weakest from nonsynonymous sites of MCS. Interestingly, including third positions of MCS increases node support to 98%. The Western Nearctic *Androprosopa* clade is comprised of three well-supported subclades encompassing six designated species groups (Fig. 2) (Pivar et al. 2018). The *An. elnora* group, comprised of *An. elnora* (Arnaud & Boussy) and *An. ericfisheri*, was recovered with absolute support and consistently placed as the sister group to the remaining western species, with node support ranging from 82 to 100%. The remaining western Nearctic species comprise two maximally

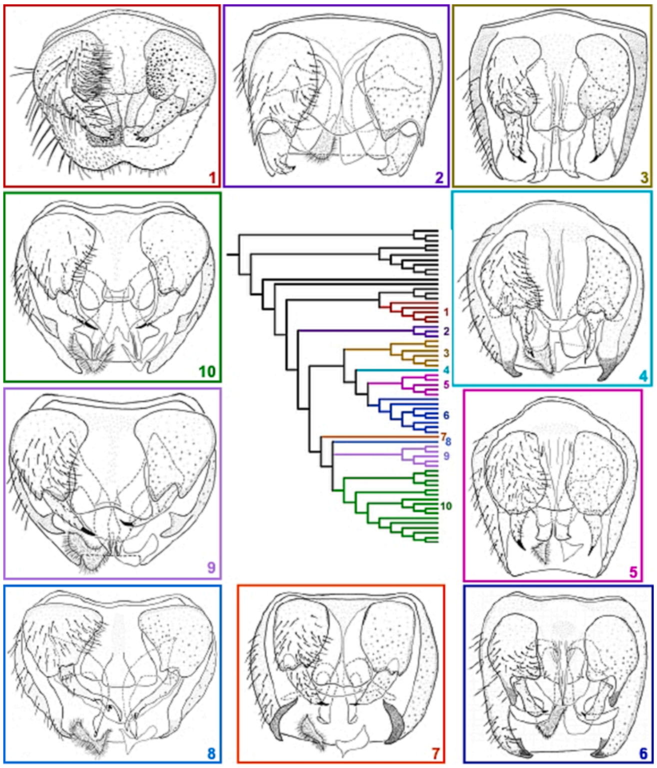


Fig. 2. Maximum likelihood tree (topology only, with *Androprosopa santaclaraensis* dropped into polytomy) with male terminalia (ventral view) of the nominal species for each currently recognized Nearctic species group provided: 1, *An. americana*; 2, *An. elnora*; 3, *An. brothersi*; 4, *An. coloradensis*; 5, *An. becca*; 6, *An. anolo*; 7, *An. chandlerorum*; 8, *An. santaclaraensis*; 9, *An. buckae*; 10, *An. gillespieae*. Illustration of *An. americana* reproduced from Sinclair (1996); remaining illustrations reproduced from Pivar et al. (2018).

supported (i.e., NS = 100% across all analyses) sister clades as follows: *An. brothersi* group (*brothersi* (Arnaud & Boussy), *confracta* (Arnaud & Boussy), *melanderi* + [*An. coloradensis* (Arnaud & Boussy) + [*An. becca* group (*becca* (Arnaud & Boussy), *lindsayorum* (Arnaud & Boussy)) + *An. anolo* group (*anolo*, *falciformis* (Arnaud & Boussy), *fusca* (Garrett), *rainierensis* Pivar & Sinclair, *waha* (Arnaud & Boussy), *zelmae* (Arnaud & Boussy)]] forming the sister group to *An. chandlerorum* (Arnaud & Boussy) + [*An. buckae* group (*buckae*, *idahoensis* (Arnaud & Boussy)) + *An. gillespieae* group (*apache* Pivar & Moulton, *arnaudi* Pivar, *gillespieae*, *palouse* (Arnaud & Boussy), *schmidiana* (Arnaud & Boussy))]. The *An. brothersi* and *becca* groups had robust node support in all analyses. *Androprosopa coloradensis* was strongly supported as the sister to the *An. becca* and *An. anolo* groups in all analyses except that of all MCS sites, where it weakly paired (66%) with the *anolo* group. Fig. 2 depicts the male terminalia of ten species that encompass the range of structural diversity observed among Nearctic *Androprosopa*.

Androprosopa chandlerorum was invariably recovered with maximum node support as sister to the clade comprised of *An. santaclaraensis* and the *An. buckae* and *An. gillespieae* groups. Placement of *An. santaclaraensis* was the most contentious among all *Androprosopa* species. It was weakly recovered as sister to the *An. buckae* group in the analysis of all sites (NS = 40%) and all MCS sites (NS = 52%) and as sister to the *An. gillespieae* group in the analysis of all nonsynonymous sites (NS = 53%) and all BZF sites (NS = 62%). It was strongly placed as the sister to the *An. buckae* and *An. gillespieae* groups in the analysis of MCS nonsynonymous sites (NS = 100%) and as sister to only one (*An. palouse*-BC) of four exemplars included for this species in the analysis of BZF nonsynonymous sites (NS = 86%). This topological uncertainty coupled with a lack of compelling morphological evidence, resulted in our relegation of *An. santaclaraensis* into an unresolved polytomy with the *An. buckae* and *An. gillespieae* groups. Relationships between multiple exemplars of *An. gillespieae* and *An. schmidiana* were the most problematic among analyses. In fact, the total evidence tree is the only one in which these two species were recovered as monophyletic, weakly so with 41% and 72% bootstrap scores, respectively.

4. Discussion

4.1. *Eastern Nearctic* Androprosopa *and Palearctic thaumaleids*

The eastern Nearctic species were strongly supported as a monophyletic clade, with *An. americana* + *An. vaillantiana* placed as sister species, and *An. thornburghae* as sister to them in all analyses. The *An. americana* + *An. vaillantiana* grouping is supported by the lack of posterolateral processes on the epandrium in males (Fig. 2.1) and stout gonostyli, shorter than length of gonocoxae, whereas *An. thornburghae* has posterolateral processes on the male epandrium and the gonostyli are narrow, distinctly tapered, longer than the gonocoxites. Monophyly of the Eastern Nearctic *Androprosopa* species is supported on the basis of the following synapomorphies: larva lacking conical protuberances on the head capsule (Sinclair 1992; Sinclair 1996) and gonostyli of the male terminalia with blunt, apical spines and inner face of gonocoxites clothed with dense setae (Sinclair 1996).

The Eastern Nearctic *Androprosopa* clade was robustly supported as the sister group to a clade comprised of all three European species sampled, the latter arranged as follows: *An. larvata* as sister to *An. striata* plus the non-congener *Thaumalea testacea*. *Thaumalea* was intended to be the proximal outgroup in this study due to morphological similarities with *Androprosopa*. Its placement here, rendering *Androprosopa* as paraphyletic, is intriguing but not unprecedented. Haubrock et al. (2017) sequenced cytochrome oxidase I and the wingless gene from an array of European thaumaleid species and observed a polyphyletic *Androprosopa*, with *An. algira* (Vaillant), treated therein as a member of *Protothaumalea*, nested within *Thaumalea*. A large-scale study employing additional markers and Palearctic species of both genera is necessary to further assess relationships among species of these two genera.

4.2. *Western Nearctic clade*

A monophyletic western Nearctic clade was consistently recovered with strong support. The western species are characterized by having long, pointed, apical gonostylar claw-like spines, and, for the most part, less dense setae on the inner face of the gonocoxites, contrasting with the character states for the eastern Nearctic species of *Androprosopa.* Six distinct species groups hypothesized by Pivar et al. (2018) were consistently recovered with strong support for each analysis. The six groups belong to three distinct clades: (1) the *An. elnora* group; (2) the cylindrical/spherical epandrium group; and (3) the quadrate epandrium group (Fig. 1).

4.2.1. An. elnora *group (Clade 1)*

The *An. elnora* species group consists of *An. elnora* + *An. ericfisheri* and was always recovered with maximum node support. Each analysis found this group to be the sister to the remaining western Nearctic clades. These species are both yellow-orange in colour (especially evident in fresh material) (Pivar et al. 2018, fig.), contrasting with the remaining Nearctic fauna, which tend toward brown and black. Another character shared between both species is the strongly tapered and anteriorly recurved gonostyli, most readily observed in lateral view (Pivar et al. 2018, figs 13, 14) (Fig. 2.2). They have been collected in cascading streams and creeks in the Pacific Northwest, particularly in the Columbia Mountains and Plateau.

4.2.2. *Cylindrical/spherical epandrium group (Clade 2)*

This group consists of species that have reniform or "bean-shaped" gonocoxites. Species in this group lack thin, posterolateral epandrial processes (except *An. coloradensis*), and the apical epandrial process is at most, a small tooth-like projection. This clade comprises three species groups: (1) *An. brothersi* group; (2) *An. becca* group; and (3) *An. anolo* group.

4.2.2.1. An. brothersi *group.* The *An. brothersi* group was consistently placed as sister to the remaining species of Clade 2 with strong support. This species groups comprises three species: *An. brothersi, An. confracta*, and *An. melanderi.* This group is defined on the basis of the parameres being curved laterally at the apex, usually with some degree of serration on the outer margin (Fig. 2.3). In addition, the gonostyli are long and gently tapered and the epandrium has four small tooth-like projections along the apical margin (Pivar et al. 2018, figs 54, 56). Members of this group are most commonly collected from streams and more rarely from seepages.

4.2.2.2. An. coloradensis *group.* Males of *An. coloradensis* exhibit unique genitalic morphology, with very narrow gonostyli and broadly flattened parameres (Fig. 2.4). Pivar et al. (2018) suggested that it be included in the *An. becca* group, primarily on the basis of the shape of the paramere; however, the molecular analyses herein strongly support its position as sister to the *An. becca* + *An. anolo* groups. *Androprosopa coloradensis* has the most developed posterolateral epandrial process in the cylindrical/spherical epandrium group.

4.2.2.3. An. becca *group.* Molecular data strongly support the monophyly of a clade containing *Androprosopa becca*, *An. lindsayorum*, and *An. sonorensis*. Although *An. sonorensis* could not be included in this study, characters of the male terminalia definitively place it within this group. All members share the derived conditions of parameres possessing a subapical process and gonostyli bearing large, apical claw-like seta(e) (Pivar et al. 2018, figs. 43, 44) (Fig. 2.5). *Androprosopa becca* has a single apical gonostylar seta that is the longest of all the Nearctic thaumaleid species. *Androprosopa lindsayorum* possesses a cluster of 3–4 long apical setae per gonostylus, much like *An. sonorensis*. Due to its resemblance to *An. lindsayorum*, *An. sonorensis* is likely its sister species, with *An. becca* sister to that pair. This is a commonly encountered group found in a variety of madicolous habitats, mainly in the Pacific Northwest. *Androprosopa sonorensis* is only known from the type locality in Sonora, Mexico, and collections in southern California and Arizona failed to recover it.

4.2.2.4. An. anolo *group.* This well supported group was recovered in all analyses and consists of the following species: *An. anolo*, *An. falciformis*, *An. fusca*, *An. rainierensis*, *An. waha* and *An. zelmae*. Members of this species group possess atypical gonostyli, ones with medial ornamentation, as in *An. rainierensis* and *An. falciformis*, or ones that are flattened, expanded, and almost devoid of setae, such as in *An. waha* and *An. zelmae*. All species, except *An. rainierensis*, have strongly recurved parameres, a synapomorphy linking them to the exclusion of their putative sister species, *An. rainierensis* (Fig. 2.6). The prolongation of the outer posterior corner of the gonocoxite serves as a synapomorphy grouping *An. anolo*, *An. fusca*, *An. waha* and *An. zelmae* to the exclusion of their putative sister species, *An. falciformis*. The *An. anolo* species group is generally restricted to the Pacific Northwest, mainly throughout the Cascade Range and the Columbia Plateau, although *An. waha* has been collected as far south as San Bernardino, CA and east to Rocky Mountain National Park (Colorado) (Pivar et al. 2018). They are typically collected in cascading mountain streams and creeks, although *An. anolo* has also been collected from roadcut seeps.

4.2.3. Quadrate epandrium group (Clade 3)

Species included in this group have a quadrate epandrium and conical gonocoxites. All species have a developed posterolateral epandrial process and a broad apical epandrial process. This clade contains four species groups: (1) *An. chandlerorum* group; (2) *An. santaclaraensis* group; (3) *An. buckae* group; and (4) *An. gillespieae* group.

4.2.3.1. An. chandlerorum *group. Androprosopa chandlerorum* was consistently recovered within Clade 3 as the sister to all remaining species in the group. Pivar et al. (2018) placed it within the *An. buckae* group based on the presence of a bifurcated posterolateral epandrial process and a bilobed gonocoxal plate (Fig. 2.7). The molecular placement at the base of Clade 3 is supported by its more spherical epandrium, reminiscent of members of Clade 2, rather than the quadrate epandria found in all other species of Clade 3, suggesting the quadrate form is synapomorphic for the group exclusive of *An. chandlerorum.* Furthermore, the gonocoxites are not conical, the apical epandrial process is absent, and the gonostyli possess four subapical setae. The confounding mix of plesiomorphic (Clade 2) and apomorphic (Clade 3) character states found in *An. chandlerorum* are perhaps close to what one might expect in a species branching between clades 2 and 3.

4.2.3.2. An. santaclaraensis *group.* The unique morphology of *An. santaclaraensis* supports placement on its own. It has unique gonostyli that appear raptorial in ventral view (Fig. 2.8) and corkscrew-like in lateral view (Pivar et al. 2018, fig. 18). The posterolateral epandrial process and hind margin of the epandrium are similar to those of *An. chandlerorum*, as both form a rather straight line in lateral view, whereas the *An. buckae* group members have a distinct C-shaped indentation formed where the posterolateral epandrial process and apical epandrial process meet along the laterad (Pivar et al. 2018). *Androprosopa buckae* group members also have a bilobed gonocoxal plate, while it is distinctly trilobed in *An. santaclaraensis* (Fig. 2.8). These morphological characters lend some credence to *An. santaclaraensis* being an intermediate species between *An. chandlerorum* and the *An. buckae* group, which is supported strongly (100%) solely by analysis of MCS nonsynonymous sites. *Androprospa santaclaraensis* also exhibits characteristics of the *An. gillespieae* group, including a trilobed gonocoxal plate. Pivar et al. (2018) hypothesized that *An. santaclaraensis* might belong to the *An. gillespieae* group on the basis of these characters, and while this relationship was recovered in three analyses (BZF and MCS nonsynonymous concatenated, BZF all codon positions and BZF nonsynonymous sites), support for this placement was generally low. This again supports the placement of this enigmatic species on its own.

4.2.3.3. An. buckae *group.* The *An. buckae* group consists of sister species *An. buckae* and *An. idahoensis*. It was consistently recovered throughout all analyses with strong support. Male genitalic synapomorphies uniting them include the bilobed distal margin of the gonocoxal plate and gonostyli bearing two dorsal subapical setae (Pivar et al. 2018) (Fig. 2.9). In addition, Arnaud & Boussy (1996) hypothesized that these two species were closely related based upon the similarities of the shape of the apex of the paramere and epandrial indentation. Gillespie et al. (1994) described immatures of several species of Nearctic *Androprosopa* and concluded that *An. buckae* and *An. idahoensis* were closely related based upon similarities in pupal abdominal chaetotaxy. Although this supports the species group, pupal descriptions from other species are too sparse to allow conveyance of convincing support. Members of this group have been collected in the Cascade, California Coastal and Sierra Nevada ranges, and east to the Columbia Plateau.

4.2.3.4. An. gillespieae *group.* The molecular data places the following species in the *An. gillespieae* group: *An. apache*, *An. arnaudi*, *An. gillespieae*, *An. palouse* and *An. schmidiana*. *Androprosopa palouse* was always recovered as sister to the remaining species. Both Arizona species, *An. apache* and *An. arnaudi* were repeatedly recovered as sister species. Monophyly of *An. gillespieae* and, to a lesser extent, *An. schmidiana* was weakly supported, being simultaneous observed only in the total evidence tree; the latter was also monophyletic in the analysis of combined nonsynonymous sites and all MCS sites. The placements of *An. gillespiae* WA and *An. schmidiana* NV appear to be the cause for this uncertainty. These results are perhaps indicative of incomplete lineage sorting or introgression. Pivar et al. (2018) hypothesized this species group based upon the presence of a simple posterolateral epandrial process, trilobed gonocoxal plate, and gonostylus with two ventrolateral subapical setae (Fig. 2.10). Based on these characters Pivar et al. (2018) included *An. zempoala* in this group (no specimens available for molecular analysis). This species group has a wide geographic range, spanning all western mountain ranges, from southern British Columbia, south to central Mexico and east to the Colorado Rockies. Members of this group have been collected in all types of madicolous habitats.

4.3. Species complexes

Several Nearctic species of *Androprosopa* exhibit wide distributions in varying habitats, perhaps indicating the possibility of species complexes or cryptic species. *Androprosopa americana*, *An. brothersi*, *An. gillespieae*, *An. palouse*, *An. schmidiana* and *An. waha* all have particularly wide distributions spanning several states, provinces and mountain ranges, however, intraspecific variation is extremely low. In fact, divergence between species within the species groups tend to be quite small (Fig. 1 inset). The *An. gillespieae* group, for example, has very few molecular changes between species; however, when coupled with morphology, clear species delimitations can be discerned.

The stem of the *An. elnora* group clade is the longest branch within the western Nearctic fauna (Fig. 1, inset), suggesting it might be the oldest of the species groups. The remaining species groups arise from very short branches comparatively, suggestive of more recent and/or rapid divergence. The quadrate epandrium group (Clade 3), in particular, appears to have radiated quickly. The branch length between eastern Nearctic species and the sampled Palearctic fauna is relatively long, perhaps reflective of under-sampling of the later.

4.4. Future considerations

This study is the first phylogenetic analysis of Nearctic *Androprosopa*. Although intended to be as comprehensive as possible, gaps remain. Molecular data is missing for *An. sonorensis* and *An. zempoala* due to our inability to obtain DNA-grade material. Their absence did not appear to adversely affect the overall integrity of the study, however, as evidenced by corroborative male genitalic structures for recovered clades. Increased genomic and taxon sampling, particularly of Palearctic species, and inclusion of morphological data, especially larval and pupal characters, would strengthen the veracity of tests of *Androprosopa* and *Thaumalea* monophyly. Thaumaleidae remain a vastly understudied family, and continued work at molecular and morphological levels will undoubtedly yield new discoveries.

CRediT authorship contribution statement

Robert J. Pivar: Investigation, Formal analysis, Data curation, Writing – original draft. **John K. Moulton:** Conceptualization, Funding acquisition, Project administration, Supervision, Resources, Formal analysis, Writing – review & editing. **Bradley J. Sinclair:** Conceptualization, Funding acquisition, Supervision, Resources, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

We thank C.J. Borkent, S.E. Brooks, J.M. Cumming, A.H. Kirk-Spriggs, R.I. Madriz, T. Saigusa and R. Wagner for providing specimens, and two anonymous reviewers for their constructive comments. The United States National Park Service provided collecting permits for Mt. Rainier National Park (Permit #: MORA-2016-SCI-0026) and Rocky Mountain National Park (ROMO-2016-SCI-0028). Cycle sequencing products were analyzed by University of Tennessee Genomics Core staff through an ABI3730 DNA Analyzer using Sequencing Analysis software 5.3.1. This research was supported by United States National Science Foundation awards DEB-0933218 and DEB-1949813 and United States Department of Agriculture National Institute of Food and Agriculture Hatch Project TEN00558.

References

Adler, P.H., D.C. Currie, Wood, D.M., 2004. Chapter 9. Phylogeny and Classification of Holarctic Black Flies. Pp. 125-160. The Black Flies (Simuliidae) of North America. Cornell University Press. 941 pp.

Arnaud Jr., P.H., Boussy, I.A., 1994. The adult Thaumaleidae (Diptera: Culicomorpha) of Western North America. Myia. 5, 41–152.

Bertone, M.A., Courtney, G.W., Wiegmann, B.M., 2008. Phylogenetic and temporal diversification of the earliest true flies (Insecta: Diptera) based on multiple nuclear genes. Syst. Entomol. 33, 668–687. https://doi.org/10.1111/j.1365-3113.2008.00437.x.

Bezzi, M., 1913. Taumaleidi (Orfnefilidi) italiani con descrizione di nuove specie. Bol. Zool. Sc. Sup. Portici. 7, 227–266.

Blaschke, J.D., Stireman III, J.O., O'Hara, J.E., Cerretti, P., Moulton, J.K., 2018. Molecular phylogenetics and piercer evolution in the bug-killing flies (Diptera: Tachinidae: Phasiinae). Syst. Entomol. 43, 218–238. https://doi.org/10.1111/syen.12272.

Borkent, A., 2012. The pupae of Culicomorpha - morphology and a new phylogenetic tree. Zootaxa. 3396, 1–98. https://doi.org/10.11646/zootaxa.3396.1.1.

Edwards, F.W., 1929. A revision of the Thaumaleidae (Dipt.). Zool. Anz. 82, 121–142.

Gillespie, J.M., Barr, W.F., Elliott, S.T., 1994. Taxonomy and biology of the immature stages of species of *Thaumalea* occurring in Idaho and California (Diptera: Thaumaleidae). Myia. 5, 153–193.

Haubrock, P.J., Kvifte, G.M., Langguth, H., Wagner, R., 2017. Glacial and postglacial species divergence and dispersal of European trickle midges (Diptera: Thaumaleidae). Arthropod Syst. Phylo. 75 (3), 523–534. https://doi.org/10.3897/asp.75.e31920.

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018. UFBoot2: Improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35, 518–522. https://doi.org/10.1093/molbev/msx281.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods. 14, 587–589. https://doi.org/10.1038/nmeth.4285.

Mackie, G.L., 2004. Applied Aquatic Ecosystem Concepts, Second Edition. Kendall Hunt Publishing, Dubuque, Iowa, p. 374.

Maddison, W.P., Maddison, D.R., 2011. Mesquite: a modular system for evolutionary analysis. Accessed 24 August 2018 Version 2, 75. http://mesquiteproject.org.

Maddison, D.R., Swofford, D.L., Maddison, W.P., 1997. NEXUS: An extensible file format for systematic information. Syst. Biol. 46, 590–621. https://doi.org/10.1093/sysbio/46.4.590.

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. Mol. Biol. Evol. 37, 1530–1534. https://doi.org/10.1093/molbev/msaa015.

Moulton, J.K., 2000. Molecular sequence data resolves basal divergences within Simuliidae (Diptera). Syst. Entomol. 25, 95–113. https://doi.org/10.1046/j.1365-3113.2000.00097.x.

Moulton, J.K., Adler, P.H., 1995. Revision of the *Simulium jenningsi* species-group (Diptera: Simuliidae). Trans. Am. Entomol. Soc. 121, 1–57.

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. Mol. Biol. Evol. 32, 268–274. https://doi.org/10.1093/molbev/msu300.

Pivar, R.J., Moulton, J.K., Sinclair, B.J., 2018. Revision of the western Nearctic *Androprosopa* (Diptera: Thaumaleidae) and descriptions of three new species. Insect Syst. Evol. 2019 (50), 346–406. https://doi.org/10.1163/1876312X-00002189.

Pivar, R.J., Sinclair, B.J., Moulton, J.K., 2021. Revision of the genus *Niphta* (Diptera: Thaumaleidae) Theischinger of South America, with descriptions of nine new species and a new immature morphotype. ZooKeys. 1063, 49–104. https://doi.org/10.3897/zookeys.1063.71180.

Rogers, J., Wall, R., 1980. A mechanism for RNA splicing. Proc. Natl. Acad. Sci. U.S.A. 77, 1877–1879. https://doi.org/10.1073/pnas.77.4.1877.

Senatore, G.L., Alexander, E.A., Adler, P.H., Moulton, J.K., 2014. Molecular systematics of the *Simulium jenningsi* species group (Diptera: Simuliidae), with three new fast-evolving nuclear genes for phylogenetic inference. Mol. Phylogenet. Evol. 75, 138–148. https://doi.org/10.1016/j.ympev.2014.02.018.

Sinclair, B.J., 1992. A new species of *Trichothaumalea* (Diptera: Thaumaleidae) from eastern North America and a discussion of male genitalic homologies. Can. Entomol. 124, 491–499. https://doi.org/10.4039/Ent124491-3.

Sinclair, B.J., 1996. A review of the Thaumaleidae (Diptera: Culicomorpha) of eastern North America, including a redefinition of the genus *Androprosopa* Mik. Entomol. Scand. 27, 361–376.

Sinclair, B.J., 2000. Immature stages of Australian *Austrothaumalea* Tonnoir and *Niphta* Theischinger (Diptera: Thaumaleidae). Aust. J. Entomol. 39, 171–176. https://doi.org/10.1046/j.1440-6055.2000.00162.x.

Sinclair, B.J., Marshall, S.A., 1987. The madicolous fauna in southern Ontario. Proc. Entomol. Soc. Ontario. 1986 (117), 9–14.

Sinclair, B.J., Pivar, R.J., Arnaud Jr., P.H., 2022. World catalogue of the family Thaumaleidae (Diptera: Culicomorpha). Zootaxa. 5225, 1–66. https://doi.org/10.11646/zootaxa.5225.1.1.

Sinclair, B.J., Saigusa, T., 2002. A new species of the seepage midge genus *Trichothaumalea* Edwards from Japan. Insect Syst. Evol. 33, 175–184.

Sinclair, B.J., Stuckenberg, B.R., 1995. Review of the Thaumaleidae (Diptera) of South Africa. Ann. Natal Mus. 36, 209–214.

Vaillant, F., 1956. Recherches sur la faune madicole (hygropétriques s.l.) de France, de Corse et d'Afrique du Nord. Ann. Mus. Hist. Nat. Paris. Nouvelle série A, Zoologie 11, 258 pp.

Vaillant, F., 1977. Les Diptères Thaumaleidae d'Europe. Ann. Soc. Entomol. Fr. 13 (4), 695–710. https://doi.org/10.1080/21686351.1977.12278667.

Vaillant, F., 1978. Les diptères Thaumaleidae d'Europe. Annls. Soc. ent. Fr. (N.S.) (1977), 13, 695-710.

Wagner, R., 2002. Insecta: Diptera: Thaumaleidae. Süßwasserfauna von Mitteleuropa. 21 (11), 39-110.

Wheeler, T.J., Kececioglu, J.D., 2007. Multiple alignments by aligning alignments. Bioinformatics. 23, 559–568. https://doi.org/10.1093/bioinformatics/btm226.

Winkler, I.S., Blaschke, J.D., Davis, J.D., Stireman III, J.O., O'Hara, J.E., Cerretti, P., Moulton, J.K., 2015. Explosive radiation or uninformative genes? Origin and early diversification of tachinid flies (Diptera: Tachinidae). Mol. Phylogenet. Evol. 88, 38–54. https://doi.org/10.1016/j.ympev.2015.03.021.