KATELYN M. MULLEN (*) (°) - SCOTT A. SCHNEIDER (†) (#) - BENJAMIN B. NORMARK (*) (†)

# NEW SINGLE-COPY NUCLEAR GENES FOR USE IN SCALE INSECT SYSTEMATICS ([1])

(*) *Department of Biology and* † *Graduate Program in Ecology and Evolutionary Biology, University of Massachusetts Amherst, Amherst, MA 01003, USA*
(#) *present address:* USDA, Agricultural Research Service, Henry A. Wallace Beltsville Agricultural Research Center, *Systematic Entomology Laboratory, 10300 Baltimore Avenue, Beltsville, MD, 20705, USA*
(°) *Corresponding author: present address: Department of Chemistry and Chemical Biology, Harvard University.* kmullen@fas.harvard.edu

Mullen K.M., Schneider S.A., Normark B.B. – New single-copy nuclear genes for use in scale insect systematics.

Despite the advent of next-generation sequencing, the polymerase chain reaction (PCR) and Sanger sequencing remain useful tools for molecular identification and systematics. To date, molecular systematics of scale insects has been constrained by the paucity of loci that researchers have been able to amplify with available PCR primers. Due to the rapid molecular evolution of scale insects, "universal" primers, and even primers developed for their sister taxon the Aphidoidea, typically fail. We used transcriptome data for two diaspidids, *Acutaspis umbonifera* (Newstead) and *Chrysomphalus aonidum* (Linnaeus), together with a published aphid genome, to design novel PCR primer sets for scale insects. Our primers amplify fragments of eight single-copy genes: ATP-dependent RNA helicase (DHX8), translation initiation factor 5 (IF5X1), DNA replication licensing factor (Mcm2), double-strand break repair protein (MRE11A), serine/threonine-protein phosphatase (PPP1CB), DNA-directed RNA polymerase II (RNApII), ribonucleoside-diphosphate reductase (RRM1), signal recognition particle receptor (SRPα), neuronal PAS domain-containing protein 4 (NPAS4), and cleft lip and palate transmembrane protein 1 (TP1). Here we report the results of tests of amplification success and phylogenetic utility of these primer sets across the Diaspididae and nine other families of Coccomorpha.

KEY WORDS: molecular systematics, PCR primers, Coccomorpha, phylogenetics.

## INTRODUCTION

Although whole-genome sequencing is becoming increasingly feasible in many systems, PCR and Sanger sequencing remain more-affordable workhorses for molecular systematics. However, for scale insects there are few loci for which PCR primer sets are available. Primer sequences have been published for only 7 nuclear loci: the large and small ribosomal subunits (18S, 28S), elongation factor 1-alpha (EF-1α), carbamoyl-phosphate synthetase (CAD), cytochrome oxidase I (COI) and cytochrome oxidase II (COII), triosphosphate isomerase, and dynamin (GULLAN and COOK, 2007; MOULTON and WIEGMANN, 2004; HARDY, 2007; ANDERSEN *et al.,* 2010); and of these, only a subset can be amplified for any given taxon, with only the ribosomal genes being consistently amplifiable. One of the reasons why molecular systematics of scale insects has been challenging is that there is a very long branch separating them from their sister taxon – the aphids – indicating that a great deal of evolutionary change has occurred over a relatively short period of time (COOK *et al.*, 2002). As a result of this divergence, primers that work for aphids and most other insects do not work well for scale insects. Using transcriptomes from two armored scale insects and an aphid, we designed primer sets for 10 loci, in the hope that they would be useful primarily in the Diaspididae and additionally across the Coccomorpha.

---

## MATERIALS AND METHODS

PRIMER DESIGN

Two armored scale insect species, *Acutaspis umbonifera* (Newstead) and *Chrysomphalus aonidum* (Linnaeus), were collected from a greenhouse in Amherst, Massachusetts, and sent to Kevin Johnson at the University of Illinois Urbana-Champaign for transcriptome sequencing. Transcriptome data were filtered using the pipeline developed by GRANT and KATZ (2014) to only include contigs that were common to the two species. Primers for each gene were designed using the Primer3 software package (UNTERGASSER *et al.,* 2012). *Acyrthosiphon pisum* transcriptome data were also used in the primer design process to determine whether primer sets could plausibly work for a wider range of taxa. These sequences were further aligned against a broad genomic dataset encompassing eukaryotes, archaea, and bacteria (GRANT and KATZ, 2014), resulting in approximately 600 candidate genes available for primer set design. Primers were designed to optimize PCR performance (APTE and DANIEL, 2009) and the best candidate primer pairs for 10 loci were tested on a set of scale insect taxa.

TAXONOMIC SAMPLING

PCR was attempted for 15 different families within the Coccomorpha for which DNA was available as of September 2014. We focused primarily on the Diaspididae because a majority of the specimens available to us belong to this family. Representatives from the Aphididae, Asterolecaniidae, Coccidae, Conchaspididae, Cryptococcidae, Eriococcidae, Kerriidae, Monophlebidae, Phoe-

nicococcidae, Pseudococcidae and Stictococcidae were included as well. Any genes that were successfully sequenced for *Acutaspis umbonifera* were also attempted for 24 additional species within the Diaspididae. Amplification was attempted for 76 specimens (Table 1). The slide-mounted cuticles of all the specimens are deposited in the University of Massachusetts Insect Collection.

PCR AND SEQUENCING

DNA preparations of *A. umbonifera* were used to run initial gradient tests, to determine the ideal annealing temperature(s) for each primer set. With every PCR, this species was also used as a positive control to aid in interpreting the success of each reaction on the resulting gels. Primer sets and their corresponding PCR protocols are listed in Table 2. Nine gene fragments were used for molecular phylogenetic analysis: translation initiation factor 5 (IF5X1), 441 base pairs; DNA replication licensing factor (Mcm2), 410 bp; double-strand break repair protein (MRE11A), 699 bp; serine/threonine-protein phosphatase (PPP1CB), 636 bp; DNA-directed RNA polymerase II (RNApII), 657 bp; ribonucleoside-diphosphate reductase (RRM1), 863 bp; signal recognition particle receptor (SRPα), 333 bp; cleft lip and palate transmembrane protein 1 (TP1), 473 bp; ATP-dependent RNA helicase (DHX8), multiple combinations of primers with different product sizes; and neuronal PAS domain-containing protein 4 (NPAS4), multiple combinations of primers with different product sizes.

Either Takara Ex Taq® hot-start polymerase (Clontech Laboratories, Inc., Mountain View, California) or GoTaq® G2 hot-start polymerase (Promega Corporation, Madison, Wisconsin) was used for standard PCR amplification. PCR protocols are listed in Table 1. The presence and size of PCR products was determined using SYBR® Safe (Life Technologies, Carlsbad, California) ultraviolet stain and 1.5% agarose gel electrophoresis. PCR products were treated with Exonuclease I and Shrimp Alkaline Phosphatase (Affymetrix, Santa Clara, California) at 37°C for 25 minutes, followed by a cycle at 80°C for 15 minutes to denature any proteins. Purified PCR products were sent to the Genomics Resource Laboratory at the University of Massachusetts Amherst for Sanger sequencing using an ABI

*Table 1* – Species for which amplification was attempted.

| Prep number | Species | Family | Amplification Success |
|---|---|---|---|
| D0991A | undetermined sp. | Aphididae | yes |
| D0991B | undetermined sp. | Aphididae | no |
| D4189A | undetermined sp. | Aleyrodidae | no |
| D2095A | *Planchonia fimbriata* (Boyer de Fonscolombe) | Asterolecaniidae | yes |
| D4195A | undetermined sp. | Asterolecaniidae | yes |
| D0320B | *Bambusaspis miliaris* (Boisduval) | Coccidae | yes |
| D0328A | *Ceroplastes rubens* Maskell | Coccidae | yes |
| D3008A | *Conchaspis* near *lepagei* | Conchaspididae | no |
| D4355A | *Conchaspis capensis* (Linnaeus) | Conchaspididae | no |
| D0583F | *Cryptococcus fagisuga* Lindinger | Cryptococcidae | yes |
| D3089A | *Acutaspis reniformis* (Cockerell) | Diaspididae | yes |
| D5045A | | | |
| D5045B | | | |
| D5045C | *Acutaspis umbonifera* (Newstead) | Diaspididae | yes |
| D5045D | | | |
| D5045E | | | |
| D5045F | | | |
| D3400B | *Aonidia lauri* (Bouché) | Diaspididae | no |
| D2052A | *Aonidia paradoxa* Lindinger | Diaspididae | yes |
| D2084A | | | |
| D3286A | *Aonidiella inornata* McKenzie | Diaspididae | yes |
| D3189A | *Aspidaspis arctostaphyli* (Cockerell & Robbins) | Diaspididae | yes |
| D2727A | *Aspidaspis densiflorae* (Bremner) | Diaspididae | yes |
| D0955A | *Aspidiella sacchari* (Cockerell) | Diaspididae | yes |
| D0703A | *Aspidiotus destructor* Signoret | Diaspididae | yes |
| D3045C | | | |
| D3519C | *Aspidiotus elaeidis* Marchal | Diaspididae | yes |
| D3536A | *Aspidiotus fularum* Balachowsky | Diaspididae | yes |
| D0595A | *Aspidiotus hedericola* Leonardi | Diaspididae | yes |
| D0620A | | | |
| D0103B | *Aspidiotus nerii* Bouché | Diaspididae | yes |
| D1876E | *Affirmaspis cederbergensis* Schneider | Diaspididae | yes |
| D1540B | *Chionaspis pinifoliae* (Fitch) | Diaspididae | no |
| D1538A | *Chionaspis pinifoliae* (Fitch) | Diaspididae | no |

*Continued Table 1*

| Prep number | Species | Family | Amplification Success |
|---|---|---|---|
| D1021A | *Chrysomphalus aonidum* (Linnaeus) | Diaspididae | no |
| D2858B | *Chrysomphalus aonidum* (Linnaeus) | Diaspididae | yes |
| D5044A | *Chrysomphalus aonidum* (Linnaeus) | Diaspididae | yes |
| D5067A | *Davidsonaspis aguacatae* (Evans, Watson & Miller) | Diaspididae | no |
| D2706A | *Diaspidiotus aesculi* (Johnson) | Diaspididae | yes |
| D1076A | *Diaspidiotus ancylus* (Putnam) | Diaspididae | yes |
| D1944A | | | |
| D1108E | *Diaspidiotus forbesi* (Johnson) | Diaspididae | yes |
| D1106E | *Diaspidiotus liquidambaris* (Kotinsky) | Diaspididae | yes |
| D2756B | *Diaspidiotus sulci* (Balachowsky) | Diaspididae | yes |
| D1401A | *Dynaspidiotus britannicus* (Newstead) | Diaspididae | yes |
| D3035A | *Hemiberlesia cyanophylli* (Signoret) | Diaspididae | yes |
| D2752A | *Lepidosaphes ulmi* (Linnaeus) | Diaspididae | no |
| D0251D | *Melanaspis obscura* (Comstock) | Diaspididae | yes |
| D1984B | *Melanaspis smilacis* (Comstock) | Diaspididae | yes |
| D1888B | *Melanaspis* sp. | Diaspididae | no |
| D3559A | *Morganella conspicua* (Brain) | Diaspididae | yes |
| D3561A | *Morganella conspicua* (Brain) | Diaspididae | yes |
| D3599A | *Morganella conspicua* (Brain) | Diaspididae | no |
| D3619A | *Morganella conspicua* (Brain) | Diaspididae | yes |
| D3146A | *Oceanaspidiotus spinosus* (Comstock) | Diaspididae | yes |
| D4898A | *Octaspidiotus stauntoniae* (Takahashi) | Diaspididae | no |
| D0281E | *Prodigiaspis* undescribed sp. | Diaspididae | yes |
| D3642A | *Pseudaulacaspis miyakoensis* (Kuwana) | Diaspididae | no |
| D3633A | *Selenaspidus albus* McKenzie | Diaspididae | no |
| D3085A | *Selenaspidus articulatus* (Morgan) | Diaspididae | yes |
| D1168E | undetermined sp. | Diaspididae | no |
| D0037A | *Eriococcus adenostornae* Ehrhorn | Eriococcidae | yes |
| D0074D | *Pseudochermes fraxini* (Kaltenbach) | Eriococcidae | yes |
| D0787A | *Ovaticoccus californicus* McKenzie | Eriococcidae | yes |
| D0333E | *Paratachardina pseudolobata* Kondo & Gullen | Kerriidae | yes |
| D0227C | *Crypticerya brasiliensis* (Hempel) | Monophlebidae | yes |
| D3269B | *Icerya purchasi* Maskell | Monophlebidae | yes |
| D1097C | undetermined sp. | Ortheziidae | no |
| D4359B | *Phoenicococcus marlatti* Cockerell | Phoenicococcidae | yes |
| D0588A | *Pseudococcus longispinus* Targioni Tozzetti | Pseudococcidae | yes |
| D1142A | *Planococcus citri* Risso | Pseudococcidae | yes |
| D4239A | undetermined sp. | Pseudococcidae | no |
| D1894C | *Myrmarachne formicaria* (De Geer) | Salticidae | yes |
| D3851A | *Parastictococcus* sp. | Stictococcidae | no |
| D3854A | *Stictococcus* sp. | Stictococcidae | yes |
| D3856A | *Stictococcus vayssieri* Richard | Stictococcidae | yes |

Model 3130XL sequencer (Life Technologies, Carlsbad, California). For some genes where multiple products were amplified and the resulting sequences could not be aligned, internal sequencing primers were designed (Table 1). These PCR products were sequenced again with these internal sequencing primers.

The resulting DNA sequences were edited and aligned in Geneious 6.1.8 (Biomatters Limited, Auckland, New Zealand). Edited sequences were imported into Mesquite 3.04 (Maddison and Maddison, 2015) and aligned with MUSCLE (Edgar 2004) for each locus. Alignments were further processed in PASTA 1.6.4 (Mirarab *et al*., 2014). The default settings were maintained, using MAFFT as the aligner tool, MUSCLE as the merger, FASTTREE as the tree estimator, and GTR+G20 as the model. These settings were applied to SRPα, Mcm2, MRE11A, and RNApII for three iterations of tree estimation and re-alignment. We decided to focus our phylogenetic analyses on these four loci because they had the best sequencing success for *A. umbonifera*. The default job settings were also maintained, with the maximum subproblem set to 50% and decomposition set to centroid.

PHYLOGENETIC ANALYSES

Three concatenated datasets were generated for phylogenetic analyses. The complete dataset contained all 42 taxa (30 species). Separate analyses were conducted for the new loci (RNApII, MRE11A, Mcm2, SRPα) and the commonly used loci (28S, CAD, EF-1α), which contained the same 24 taxa (22 species). Both of the latter consensus

*Table 2* – PCR primers and protocols.

| Locus | Forward Primer | Reverse Primer | Annealing Temperature Protocol | Families for which amplification was successful |
|---|---|---|---|---|
| SRPα | SRPα_530F<br>5' - ATG CRG CTG GAA TTG CYA TGG AAG C - 3'<br><br>Internal sequencing primer SRPα_548intF<br>5' - GGA ATT GCC ATG GAA GCK AT - 3' | SRPα_863R<br>5' - CCT GTG ATG TAT GTC ATY GAW ATA GC - 3' | 58→50°C (-1°/4 cycles) + 3 cycles at 50°C for 1 minute | Aphididae, Asterolecaniidae, Coccidae, Cryptococcidae, Diaspididae, Eriococcidae, Phoenicococcidae, Pseudococcidae |
| RNApII | RNApII_2519F<br>5' - ATA CGG CTG TGA AAA CTG CTG A - 3' | RNApII_3176R<br>5' - CGA TCA ACC ATT CAA ACG CTT C - 3' | 57→51°C (-1°/3 cycles) + 15 cycles at 51°C for 1.5 minutes | Coccidae, Diaspididae, Pseudococcidae |
| Mcm2 | Mcm2_1333F<br>5' - CCA GGT GAS AAA CAT AAA MTY CGT GG - 3' | Mcm2_1743R<br>5' - RAA AGT CAT AGC YGG ATC ATA SCG WCC - 3' | 58→48°C (-1°/3 cycles) + 5 cycles at 48°C for 1 minute | Aphididae*, Diaspididae |
| MRE11A | MRE11A_42F<br>5' - CGT AGC CAC TGA TAT TCA TCT MGG - 3' | MRE11A_741R<br>5' - ACA CTC GTG TTC GTG ACC C - 3' | 57→51°C (-1°/3 cycles) + 15 cycles at 51°C for 1.5 minutes | Diaspididae, Phoenicococcidae |
| NPAS4 | NPAS4_19F<br>5' - TTC TTG ATG ATGATG ACG CA - 3' | NPAS4_802R<br>5' - GRA GTT TAC TYT GCA CCA TGT - 3' | 35 cycles at 48°C for 1.5 minutes | Asterolecaniidae, Coccidae, Diaspididae, Eriococcidae*, Monophlebidae |
| | NPAS4_68F<br>5' - AC GCC GCC GAA TAC TTG G - 3' | NPAS4_670R<br>5' - CT TGT AGA ACG CAR TGA ATC CA - 3' | 54→52°C for 1 minute | Stictococcidae* |
| IF5X1 | IF5X1_29F<br>5' - TCA GYG ATG CRT TCT AYC GR - 3' | IF5X1_469R<br>5' - ATG ATC CYT GAG CKG CWG G - 3' | 57→51 (-1°/3 cycles) + 15 cycles at 51°C for 1.5 minutes | Cryptococcidae, Pseudococcidae |
| | | | 40 cycles at 51°C for 1 minute | Aphididae |
| | | | 40 cycles at 53°C for 1 minute | Phoenicococcidae, Diaspididae |
| DHX8 | DHX9_2F<br>5' - GAC GGY ATG TTA YTR CGW GA - 3' | DHX8_2R<br>5' - CMG GWA CAG GHG TTG GTA AC - 3' | 57→51°C (-1°/3 cycles) + 15 cycles at 51°C for 1.5 minutes | Coccidae, Eriococcidae, Asterolecaniidae, Diaspididae |
| | | DHX8_1R<br>5' - GGT ARW GCM GAA TAY ACA GGW A - 3' | 5 cycles at 45°C, then 49→54°C (+1°/5 cycles) for 1 minute | Asterolecaniidae, Diaspididae |
| | DHX8_1041F<br>5' - AAA CAR AGR GCT GGT CGT GC - 3' | DHX8_1340R<br>5' - KGG AAA YTC TGC CAT TCG TCT TCC - 3' | 58→48°C (-1°/3 cycles) + 5 cycles at 48°C for 1 minute | Asterolecaniidae* |
| RRM1 | RRM1_2F<br>5' - GGM AAA GTT GTB GAA CGT C - 3' | RRM1_2R<br>5' - TRC AWA CWG CTA CYT CWT C - 3' | 35 cycles at 49°C for 1 minute | Coccidae, Kerriidae, Monophlebidae |
| | | | 35 cycles at 48°C for 1 minute | Asterolecaniidae, Cryptococcidae, Diaspididae, Monophlebidae, Pseudococcidae, Stictococcidae |
| PPP1CB | PPP1CB_198F<br>5' - WTT ATT TGA GTA TGG AGG CTT TCC - 3' | PPP1CB_834R<br>5' - RTC AAC AGA CAT CAT WCC ACC WGC - 3' | 35 cycles at 54°C for 1 minute | Diaspididae, Phoenicococcidae |
| TP1 | TP1_89F<br>5' - AWM YYT CGA CCG TTC ATG C - 3' | TP1_562R<br>5' - SRT CAT CWK CAT CTG CAT CC - 3' | 35 cycles at 49°C for 1 minute | Eriococcidae, Phoenicococcidae |
| | | | 5 cycles at 45°C, then 30 cycles at 52°C for 1 minute | Pseudococcidae |

* Indicates families that have not been sequenced. Families represented reflect families that have been sequenced both successfully and unsuccessfully.

trees were compared to evaluate the performance of the new gene set against the set of commonly used loci. Each new gene region was also analyzed independently. No phylogenetic analyses were performed on NPAS4, IF5X1, DHX8, RRM1, PPP1CB, or TP1 due to a lack of successful sequencing for *A. umbonifera,* coupled with limited resources. The nucleotide sequence dataset was nearly complete for the standard set of loci – one taxon is missing data for 28S and three taxa are missing data for EF-1α. Gaps in coverage are more severe across the new loci, which might reflect a need for further troubleshooting of PCR and/or sequencing protocols. The dataset for new loci is missing data for 33-45% of specimens.

Model selection analyses were run for each locus in

jModelTest 2.1.7 (DARRIBA *et al*., 2012). For SRPα, RNApII, and MRE11A, the preferred evolutionary model is the Hasegawa-Kishino-Yano model (HKY), with a proportion of invariant sites and gamma-distributed rates (HKY+I+G). For Mcm2 the preferred evolutionary model was also HKY, with a proportion of invariant sites (HKY+I). These preferred models were implemented in all subsequent analyses.

Bayesian inference using Metropolis-coupled Markov chain Monte Carlo (MC 3) methods were used in MrBayes 3.2.6 (RONQUIST and HUELSENBECK, 2003) to reconstruct a phylogeny of Coccomorpha specimens based upon DNA sequence data. Computational resources from Cyberinfrastructure for Phylogenetic Research (CIPRES) Science Gateway 3.3 (MILLER *et al.,* 2010) were used to complete these analyses. For each analysis, 2 independent runs were conducted concurrently with 4 chains each (3 hot, 1 cold). Each analysis of concatenated datasets was allowed to run for 10 million generations, sampling parameters every 1000 generations.

For the concatenated runs, stationarity was reached by 250 thousand generations as determined by visualizing the likelihood-by-generation plot, the potential scale reduction factor (PSRF $\cong$ 1.0), and the standard deviation of split frequencies ($\leq$ 0.01). The first 25% of generations were discarded as the burn-in, leaving a total of 7,501 trees from each run available for reconstruction of a majority-rule consensus tree. Consensus trees for each analysis were generated using the sumt command in MrBayes, providing branch lengths as substitutions per site and branch support values as posterior probabilities. Subsequent visual editing of the resulting trees was conducted in FigTree 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/). For independent genealogical analyses, the same methods as above were followed except that analyses were allowed to run for 5 million generations; the burn-in was also set to 25% of generations, leaving 3,751 trees from each run available for constructing each consensus tree. Genealogies were used to assess congruence of nodes on the concatenated majority-rule consensus trees for the new and standard loci.

## RESULTS

PCR AMPLIFICATION AND SEQUENCING WITH NEW PRIMERS

All ten loci were successfully amplified for the Diaspididae. Table 2 gives a list of taxa for which at least one amplification was successful, as well as the number of specimens per species. Some PCR products were not sequenced, as indicated in Table 1. The most successful primer set designed was that for SRPα, which was successfully sequenced for 20 species representing five families across the Coccomorpha. However, successful sequencing required the use of an additional forward internal sequencing primer, and nearly half of the families sequenced with this aid (Asterolecaniidae, Cryptococcidae, and Pseudococcidae) still were not sequenced successfully, so there is room for improvement. RRM1 was very successful in amplifying for nine families across the Coccomorpha, but yielded low-quality sequence data even after utilizing internal sequencing primers.

PHYLOGENETIC ANALYSES

Family-level relationships recovered from our analyses are largely inconsistent with previous studies of Coccomorpha phylogeny (COOK *et al*., 2002; GULLAN and COOK, 2007). However, reconstructing a robust phylogeny

of Coccomorpha was not a goal of this study and we hypothesize that the inconsistencies of higher-level relationships found are likely an artifact of incomplete character sampling. The careful studies of COOK *et al*. (2002) and GULLAN and COOK (2007) present a more reliable estimate of family-level relationships among the Coccomorpha.

Focusing on relationships within Diaspididae (which relates more directly to the goals of this paper), we find relationships very closely matching expectations based upon previous studies (MORSE and NORMARK, 2006; ANDERSEN *et al.*, 2010) and more recently completed work (SCHNEIDER *et al.*, in prep) (Fig. I). A strongly-supported (pp = 1.0) monophyletic clade of "core aspidiotines", *sensu* ANDERSEN *et al.* (2010), is recovered in our analysis (Fig. I). The clade representing multiple *Diaspidiotus* species and close relatives and the clade including *Aspidiotus nerii* and its relatives, are consistent with previous results (SCHNEIDER et al., in prep). We also find that the addition of new loci helps to further resolve relationships that have either proven difficult to resolve in previous efforts or have yielded different results. For example, recent phylogenetic reconstructions of Diaspididae recover *Aspidiotus hedericola* as sister to the tribe Aspidiotini (SCHNEIDER *et al*., in prep), but our results recover this species nested within a Palearctic/Afrotropical clade of *Aspidiotus* and *Selenaspidus* species with weak support (posterior probability = 0.86, Figure I). The position of the South African species *Aspidiotus elaeidis* and *A. fularum* as a part of this Afrotropical/Palearctic clade is unique to this analysis. The previously indeterminate placement of *Selenaspidus articulatus* was resolved as sister to the latter clade with weak support (posterior probability = 0.69, Fig. I).

The juxtaposed majority-rule consensus trees resulting from independent analyses of standard vs. new loci (**1** and **2** respectively – Fig. II) show how the two sets of loci are congruent in some clades and differ – either in resolution or composition – in others. Even with a higher degree of missing data, the new set of loci (**2** – Fig. II) recover the same relationships as the standard set (**1** – Fig. II) for the *Diaspidiotus* + relatives clade mentioned above. *Aspidiella sacchari*, *Oceanaspidiotus spinosus* and *Aspidiotus destructor* are recovered with the same relationships in both as well. The *Melanaspis* species are found to be more closely related to *A. nerii* + relatives in **1**, and more closely related to *Selenaspidus* in **2** (Fig. II). Take note that *Melanaspis smilacis* is missing a large amount of character data, which could account for its equivocal position in **2**(Figure II). The full dataset analysis (Fig. I) is consistent with the results of **2** (Fig. II) in this respect, showing *Melanaspis + Acutaspis* as sister to the Afro-tropical/Palearctic clade – albeit weakly supported (pp = 0.66, Fig. I). *Aspidiotus hedericola* is positioned as sister to the rest of the aspidiotine species in **1**; its position is unresolved in **2**. The resolution of *A. elaeidis* + *A. fularum* is improved in **2**, and is recovered as sister to *Selenaspidus* + *Melanaspis*.

## DISCUSSION

Our analyses suggest that SRPα, RNApII, Mcm2, and MRE11A may serve as informative genes for phylogenetic reconstruction of the Diaspididae. Great advances in our understanding of diaspidid phylogenetic relationships have been made using data from 28S, EF-1α, COI-COII, and CAD (MORSE and NORMARK 2006; ANDERSEN *et al*., 2010; SCHNEIDER *et al*., in prep).
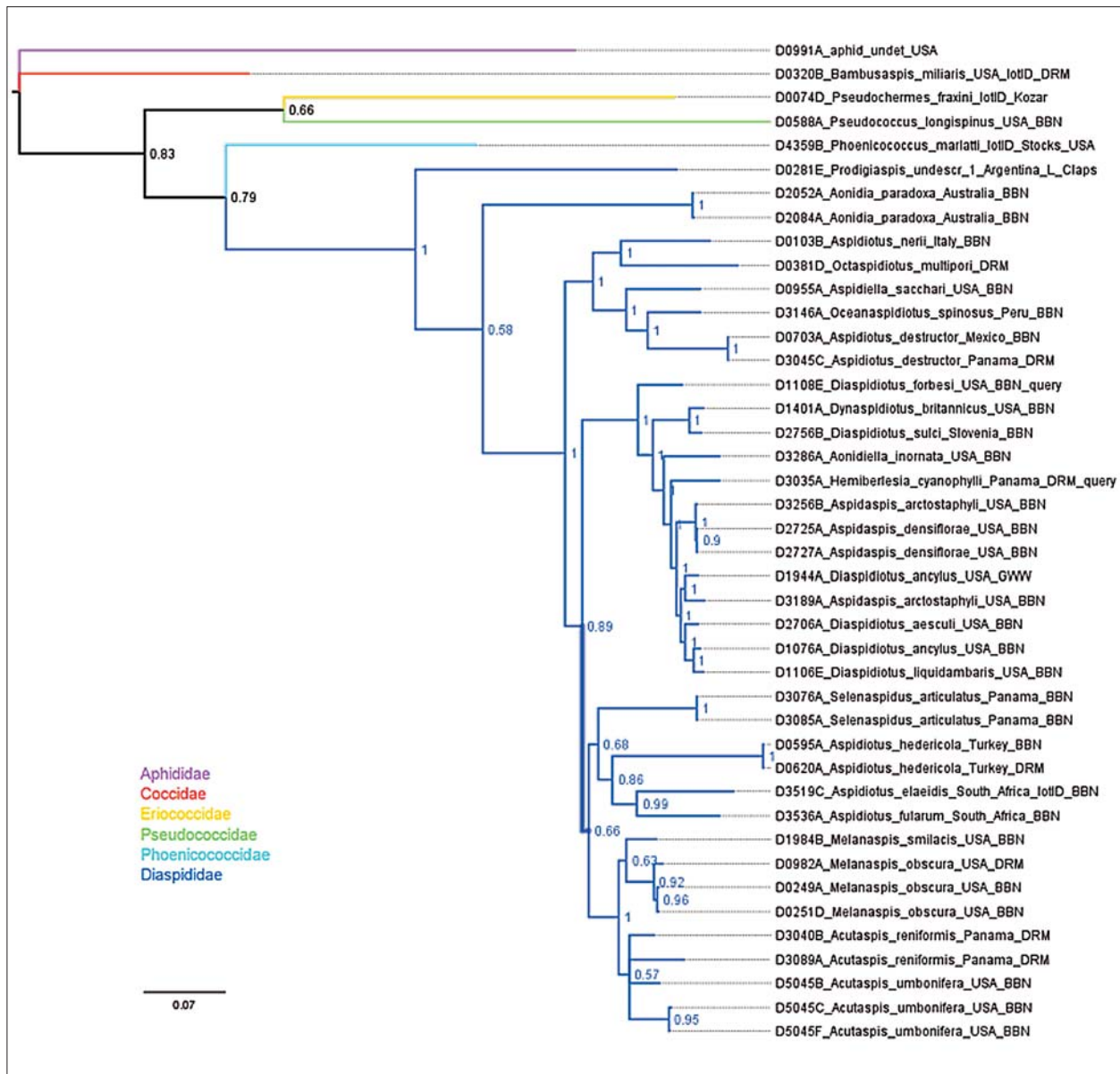
Fig. I – Phylogeny of the superfamily Coccomorpha. The majority-rule consensus tree resulting from Bayesian analysis of the concatenated dataset for seven gene regions (28S, EF-1α, CAD, SRPα, RNApII, Mcm2, MRE11A) of 42 taxa (30 species). Taxa are listed as: Identification # (e.g. D0991A)_Species name_Locality_Identifier. The specimen designated as "undescr" is an undescribed new species of *Prodigiaspis*. Branch support values are indicated as posterior probabilities. Families are color coded by branch as indicated by the legend.

Our newly developed primer sets demonstrate great potential for continuing this work. SRPα, RNApII, Mcm2, and MRE11A add new information to the phylogeny that can help address outstanding questions about relationships between diaspidids. For instance, *Aspidiotus hedericola* was recovered as part of a Palearctic/Afrotropical clade nested within Aspidiotini, rather than as sister to the rest of Aspidiotini as has been found previously (SCHNEIDER *et al.*, in prep).

In addition to the four primer sets we used to reconstruct a phylogeny of species from Coccomorpha (SRPα, RNApII, Mcm2, MRE11A), we lay the groundwork for potentially incorporating six additional loci in such analyses. The relative sequencing success of IF5X1, DHX8, TP1, PPP1CB, NPAS4, and RRM1, as well as details regarding which families each primer set worked for, are provided as a starting point for continuation of this work.
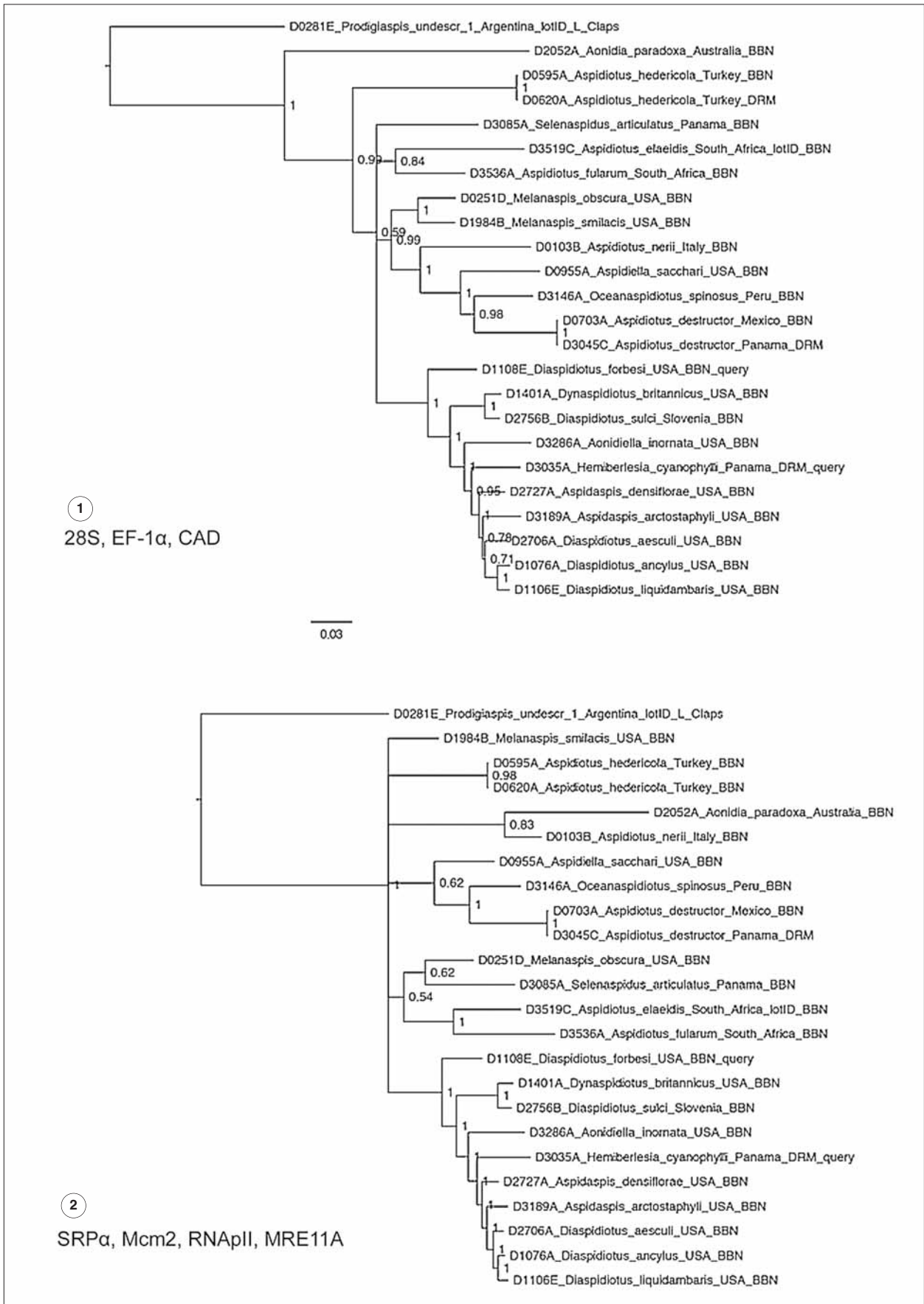
Fig. II – Majority-rule consensus trees of standard (1) vs. new (2) loci. These trees resulted from Bayesian analysis of the concatenated datasets for (1) three gene regions (28S, EF-1α, CAD) of 24 taxa (22 species) and for (2) four gene regions (SRPα, Mcm2, MRE11A, RNApII) of 34 taxa (29 species). Taxa are listed as: Identification # (e.g. D0991A)_Species name_Locality_Identifier. The specimen designated as "undescr" is an undescribed new species of *Prodigiaspis*. Branch support values are indicated as posterior probabilities.

Agricultural Experiment Station (Hatch MAS00452). Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture; USDA is an equal opportunity provider and employer.

REFERENCES

ANDERSEN J.C., WU J., GRUWELL M.E., GWIAZDOWSKI R., SANTANA S.E., FELICIANO N.M., MORSE G.E., NORMARK B.B., 2010 – *A phylogenetic analysis of armored scale insects (Hemiptera: Diaspididae), based upon nuclear, mitochondrial, and endosymbiont gene sequences* - Mol. Phylogenet. Evol., 57 (3): 992-1003.

APTE A., DANIEL S., 2009 – *PCR primer design*. Cold Spring Harbor Protocols, 3: pdb-ip65.

COOK L.G., GULLAN P.J., TRUEMAN H.E., 2002 – *A preliminary phylogeny of the scale insects (Hemiptera: Sternorrhyncha: Coccoidea) based on nuclear small-subunit ribosomal DNA*. - Mol. Phylogenet. Evol., 25 (1): 43-52.

DARRIBA D., TABOADA G.L., DOALLO R., POSADA D., 2012 – *jModelTest 2: more models, new heuristics and parallel computing.* - Nat. Methods, 9 (8): 772.

GRANT J.R., KATZ L.A., 2014 – *Building a phylogenomic pipeline for the eukaryotic tree of life - addressing deep phylogenies with genome-scale data.* - PLoS Currents Tree of Life, 1.

GULLAN P.J., COOK L.G., 2007 – *Phylogeny and higher classification of the scale insects (Hemiptera: Sternorrhyncha: Coccoidea).* - Zootaxa, 1668: 413-425.

HARDY N.B., 2007 – *Phylogenetic utility of dynamin and triose phosphate.* - Zootaxa, 1668: 413-425.

MADDISON W.P., MADDISON D.R., 2015 – *Mesquite: a modular system for evolutionary analysis - Version 3.04.* http://mesquiteproject.org.

MILLER M.A., PFEIFFER, W., SCHWARTZ, T., 2010 – *Creating the CIPRES Science Gateway for inference of large phylogenetic trees.* In: Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA pp 1-8.

MIRARAB S., NGUYEN N., WARNOW T., 2014 – *PASTA: ultra-large multiple sequence alignment.* In: International Conference on Research in Computational Molecular Biology, Springer International Publishing, Switzerland, pp. 177-191.

MORSE G.E., NORMARK B.B., 2006 – *A molecular phylogenetic study of armoured scale insects (Hemiptera: Diaspididae).* - Syst. Entomol., 31: 338-349.

MOULTON J.K., WIEGMANN B.M., 2004 – *Evolution and phylogenetic utility of CAD (rudimentary) among Mesozoic-aged Eremoneuran Diptera (Insecta).* - Mol. Phylogenet. Evol, 31: 363-378.

RONQUIST F., HUELSENBECK J.P., 2003 – *MrBayes 3: Bayesian phylogenetic inference under mixed models*. - Bioinformatics, 19(12): 1572-1574.

UNGERGASSER A., CUTCUTACHE I., KORESSAAR T., YE J., FAIRCLOTH B.C., REMM M., ROZEN S.G., 2012 – *Primer3 - new capabilities and interfaces*. - Nucleic Acids Research, 40: e115.