Biological Control 152 (2021) 104426

Contents lists available at ScienceDirect

Biological Control

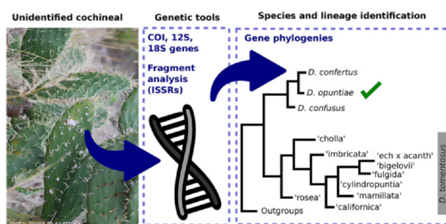
journal homepage: www.elsevier.com/locate/ybcon

Addressing the red flags in cochineal identification: The use of molecular techniques to identify cochineal insects that are used as biological control agents for invasive alien cacti

C.J.M. van Steenderen*,[1], I.D. Paterson[1], S. Edwards[2], M.D. Day[3]

Department of Zoology and Entomology, Rhodes University, Grahamstown 6140, South Africa

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Cactaceae
Dactylopiidae
DNA barcoding
ISSR

ABSTRACT

Invasive Cactaceae cause considerable damage to ecosystem function and agricultural practices around the world. The most successful biological control agents used to combat this group of weeds belong to the genus *Dactylopius* (Hemiptera: Dactylopiidae), commonly known as 'cochineal'. Effective control relies on selecting the correct species, or in some cases, the most effective intraspecific lineage, of cochineal for the target cactus species. Many of the *Dactylopius* species are so morphologically similar, and in the case of intraspecific lineages, identical, that numerous misidentifications have been made in the past. These errors have resulted in failed attempts at the biological control of some cactus species. This study aimed to generate a multi-locus genetic database to enable the accurate identification of dactylopiids. Genetic characterization was achieved through the nucleotide sequencing of three gene regions (12S rRNA, 18S rRNA, and COI) and two inter-simple sequence repeats (ISSR). Nucleotide sequences were very effective for species-level and *D. tomentosus* lineage-level identification, but could not distinguish between the two lineages within *D. opuntiae* commonly used for biological control of various *Opuntia* spp. Fragment analysis through the use of ISSRs successfully addressed this issue. This is the first time that a method has been developed that can distinguish between these two *D. opuntiae* lineages. Using the methods developed in this study, biological control practitioners can ensure that the most effective agent species and lineages are used for each cactus target weed, thus maximizing the level of control.

* Corresponding author.
E-mail address: vsteenderen@gmail.com (C.J.M. van Steenderen).
[1] Centre for Biological Control, Rhodes University, Grahamstown, 6140, South Africa.
[2] Zoology and Entomology Molecular Lab, Rhodes University, Grahamstown, 6140, South Africa.
[3] Department of Agriculture and Fisheries, GPO Box 267, Brisbane, Queensland, 4001, Australia.

https://doi.org/10.1016/j.biocontrol.2020.104426
Received 6 June 2020; Received in revised form 21 August 2020; Accepted 27 August 2020
Available online 03 September 2020
1049-9644/ © 2020 Elsevier Inc. All rights reserved.

1. Introduction

One of the major groups of invasive alien plants (IAP) in both South Africa and Australia are the Cactaceae (Klein, 2011; Paterson et al., 2011; Winston et al., 2014). Approximately 400 cactus taxa have been introduced to South Africa (Kaplan et al., 2017), of which 35 species are invasive (Novoa et al., 2015). Australia currently has 39 recorded invasive cactus taxa belonging to four genera; and is the most heavily invaded of all the countries surveyed by Novoa et al., 2015.

The most prominent biological control agents for invasive *Opuntia* and *Cylindropuntia* species are *Cactoblastis cactorum* Berg (Lepidoptera: Pyralidae) and four *Dactylopius* Costa (Hemiptera: Dactylopiidae) species; namely *D. austrinus* De Lotto, *D. ceylonicus* Green, *D. opuntiae* Cockerell, and *D. tomentosus* Lamarck (Sheehan and Potter, 2017). *Cactoblastis cactorum* has been used to control eight invasive cactus species worldwide, while the various *Dactylopius* spp. have been used to control 14 invasive cactus species across the world (Winston et al., 2014).

The Dactylopiidae family is monogeneric, comprises eleven species, is indigenous to the Americas, and is predominantly host specific to Opuntioid cacti (De Lotto, 1974; Campana et al., 2015; Van Dam et al., 2015). *Dactylopius opuntiae* and *D. tomentosus* are two species that are known to contain distinct genetic groups that display differential host specificities (Volchansky et al., 1999; Jones et al., 2015; Mathenge et al., 2015). These intraspecific groups are frequently referred to as 'biotypes' in the literature, but in agreement with the criticism raised by Downie, 2010, the term 'lineage' is used here instead. Intraspecific lineages currently used within *D. opuntiae* comprise 'ficus' and 'stricta', and those within *D. tomentosus* comprise 'bigelovii', 'californica var. parkeri', 'cholla', 'cylindropuntia sp.', 'echinocarpa x acanthocarpa', and 'imbricata' (Paterson et al., 2011; Jones et al., 2016b). It is currently impossible to distinguish between *Dactylopius* lineages using morphological characteristics (Mathenge et al., 2015; Jones et al., 2016a), raising a need to address this using genetic identification tools.

The taxonomy of the Dactylopiidae is largely understudied, and until fairly recently, only morphological characters were used to create phylogenies (Ramírez-Puebla et al., 2010). The history of this insect's taxonomy is rooted in misidentifications, which have led to more than one case of failed biological control efforts in the past (e.g. Brown et al., 1985; Volchansky et al., 1999). Many of the species are notoriously difficult to differentiate using morphological traits, even for experts (Mann, 1969; Pérez-Guerra and Kosztarab, 1992; Gullan and Kosztarab, 1997; Portillo and Vigueras, 2006). Considering the high species diversity of the Cactaceae, the probability of there being many more *Dactylopius* species and lineages in the native range is very high (Jones et al., 2016a).

Different *Dactylopius* species and lineages display different levels of population growth and impact on different Cactaceae, and so correctly distinguishing between them is fundamental to selecting the most effective agents for biological control programmes. This is where the use of genetic barcoding can be useful. The aim of this study was therefore to create a comprehensive genetic database to allow for the accurate identification of the species and intraspecific lineages within the Dactylopiidae.

2. Materials and methods

2.1. Sourcing of samples

A detailed list of all the samples used in this study appears in Table S1. Outgroup and additional ingroup sequences were obtained from GenBank (https://www.ncbi.nlm.nih.gov/genbank/), and are listed in Table S2.

We sourced species and lineages of cochineal insects used for biological control from the collections of the Department of Agriculture and Fisheries, Queensland, for Australian samples, and from the Centre for Biological Control (CBC) Cactus Mass-rearing Facility, Rhodes University, for South African samples. Other species not held in culture by either organization were field collected. Species level identifications that could not be conducted by the authors were confirmed by expert taxonomist Ian Miller at the South African National Collection of Insects (ARC-PPRI), where voucher specimens are housed (Table S3). For the two lineages of *Dactylopius opuntiae*, the 'stricta' and 'ficus' lineages, specimens were collected from cultures housed by the CBC that were known to be the correct lineage, and from field sites where the correct lineage was known to occur. The intention was to test whether specimens known to be 'stricta' or 'ficus' could be separated with this analysis. The geographic location where the samples were collected from is therefore less important than the knowledge of the source of each cochineal population as either 'stricta' or 'ficus'. 'Stricta' lineages were sourced from the CBC's 'stricta' culture, the Kruger National Park, and Saudi Arabia, where the 'stricta' lineage was released and known to have established. The 'ficus' lineage was collected from field sites in South Africa where it was known to have established, and in Namibia, where only the 'ficus' lineage has been released, and from the CBC mass-rearing facility (Table S1).

2.2. Sample preservation and DNA extraction

All specimens were preserved in 95% ethanol and stored at −18 °C. Insect tissue was digested in Buffer ATL (Qiagen©) and Proteinase K (Qiagen©), followed by a standard salt extraction procedure (Bruford et al., 1992). DNA pellets were resuspended in 100 μL AE buffer (Qiagen©) and stored at −18 °C.

2.3. Polymerase chain reaction (PCR) and sequencing

PCR protocols were adapted from the methods of Park et al. (2010), Campana et al. (2015) and Mathenge et al. (2015) (Table S4). The two COI primer pairs used in the present study (COI-A and COI-B) were checked to ascertain whether they amplify the same region of the COI gene by aligning representative sequences derived from each primer to the whole mitochondrial sequence of *Ceroplastes japonicus* Green (Coccoidea: Coccidae) (Genbank ID: MK847519.1). Although they do amplify the same COI region, it was found here that COI-A only worked for *D. confertus* De Lotto, *D. confusus* Cockerell and *D. opuntiae*; while COI-B only worked for *D. tomentosus*, as it was particularly designed by Mathenge et al., 2015 for this species. Table S5 provides the mixture reagent concentrations used for sample preparation, which was followed by amplification in a Bio-Rad T100 Thermal Cycler™ according to the protocols given in Table S6. 12S primers had a tendency to produce double bands for some samples (the target fragment plus a fragment approximately 1500 bp in size). Touchdown PCR (TD-PCR) was applied in such cases according to the protocol in Table S7. Amplified PCR products were purified, and sequenced in the forward direction in all cases by Macrogen Inc. in the Netherlands.

2.4. Sequence alignment

Chromas v2.6.4 (Technelysium Pty Ltd.) was used to view and trim (±50–70 bp at the start and end of each sequence) chromatograms. BioEdit v7.0.5.3 (Hall, 1999) was used to verify base calls against corresponding chromatograms, and to ensure that sequences were in the longest open-reading frame. MAFFT v7 (Katoh et al., 2017) was used to align sequences on the online server (https://mafft.cbrc.jp/alignment/server/).

2.5. Construction of phylogenetic trees

DAMBE v6 (Xia, 2018) was used to test substitution saturation for the protein-coding COI sequences, using the tests created by Xia et al. (2003) and Xia and Lemey (2009), at positions 1, 2, 3, and 1 & 2. The

COI sequences displayed little saturation at all codon positions (Iss < Iss.cSym value, with a p-value < 0.05 for all tests), and so additional codon models were not necessary in the construction of the COI phylogeny. jModelTest v2.1.10 (Guindon and Gascuel, 2003; Darriba et al., 2012) was used to select appropriate Bayesian information criterion (BIC) evolutionary models, where the optimal models used were GTR + G for 12S, TIM2ef + I for 18S, and HKY + G for COI. Bayesian Inference (BI) and Maximum Likelihood (ML) methods were subsequently employed to create single gene trees. Kimura-2-parameter (K2P) intra-and inter-specific distance values were recorded for all phylogenies. The congruence of the topologies of the individual gene trees for the 12S and 18S gene regions were compared using the online congruency index test 'Icong' (http://max2.ese.u-psud.fr/icong/index.help.html), developed by de Vienne et al., 2007, in order to create a concatenated phylogeny. MEGA v7 (Kumar et al., 2016) was used to create a neighbour-joining tree for each gene region, using the default settings. The pair of topologies were subsequently uploaded to the Icong server. A significant Icong p-value supported congruency (Icong = 1.72, p = 2.47×10^{-8}), and thus the concatenation of the two gene regions.

Each gene was partitioned as a whole in order to create a concatenated BI and ML phylogeny. The COI region was not included in this concatenation, because it contained only 49% of the sequences represented by the 12S and 18S datasets. A test concatenated phylogeny with all three genes did not show well-resolved clades within *D. tomentosus*.

2.5.1. Bayesian inference and maximum likelihood

MrBayes v3.2.6 (Huelsenbeck and Ronquist, 2001) was employed to create Bayesian phylogenies (random starting trees, four chains (three hot and one cold), 50 million generations). Trees were sampled every 1000 generations with a burnin of 1250. Tracer v1.7 was used to check for MCMC convergence, such that the effective sample size (ESS) was greater than 200 for all parameters (Rambaut et al., 2018).

GARLI v2.01 (Zwickl, 2006) and CONSENSE v3.69 (unrooted, consensus type Majority Rule (extended) (MRe)) (Felsenstein, 2006) were employed to create Maximum Likelihood (Felsenstein, 1981) phylogenies, where 1000 bootstrap repeats were run with two search repetitions. The CIPRES Science Gateway portal (Miller et al., 2010) was used to generate all phylogenies, and all were viewed in FigTree v1.4.3.

2.5.2. Haplotype networks in POPART

POPART v1.7 (Leigh and Bryant, 2015) was used to create 12S and COI haplotype networks for all the *D. opuntiae* and *D. confusus* in each respective phylogeny (see Figs. S1 and S3, and the sample list in Table S1) using the TCS network method, with an added 'traits block' to color haplotypes by source and/or geographic location (Clement et al., 2000). For *D. opuntiae*, samples were grouped into Uitenhage mass rearing facility (MRF), the United States of America, Australia, and 'ficus' and 'stricta' lineages collected in South Africa. *D. confusus* was divided into groups from Tucson (Arizona), Four Peaks Mountain (Arizona), Las Cruces (New Mexico), and Laredo (Texas), in the United States of America. *Dactylopius tomentosus* haplotypes were not presented here because the 12S and COI phylogenies sufficiently represented the different lineages. It was, however, used to report the number of nucleotide differences between lineages of interest in this species. The 18S gene did not provide enough intraspecific variation, and therefore was not included.

2.6. Distance-based testing of barcodes

The R 'spider' package v1.5.0 (Brown et al., 2012) was used to test barcodes in R v3.6.1 (R Core Team, 2019) in RStudio v1.2.5033 (RStudio Team, 2015) at both the species and lineage level. This entailed the use of the Threshold Identification (TID), and Best Close Match (BCM) methods. The Nearest Neighbor (NN) method was not considered further because it resulted in an abnormally high number of false positive identifications due to a nonadjustable distance threshold of 1%. This is not useful when different genes are being analysed that display different evolutionary rates. The TID threshold percentage was optimized to minimize cumulative error (false negatives + false positives). This was achieved by using the threshold optimization function in the package to test a range of threshold values from 0.0001% to 2.5% in increments of 0.005%. Pairwise distances were calculated in the R 'ape' package v5.3 (Paradis et al., 2004), using the dna.dist function and K80 substitution model. The genetic barcoding gap was calculated and plotted by calculating the largest intraspecific and the smallest interspecific genetic distance for each sequence (Meier et al., 2006), using the spider package. Scoring the accuracy of genetic barcodes followed the criteria presented by Birch et al., 2017 and in the R spider documentation (Brown et al., 2012).

2.7. Inter-simple sequence repeats (ISSRs)

2.7.1. PCR protocol, data capturing and data processing

Universal ISSR primers 809 (5'-AGA GAG AGA GAG AGA GC -3') and 826 (5'-ACA CAC ACA CAC ACA CC −3') from Primer set # 9 of the University of British Columbia Nucleic Acid Protein Service Unit were used (Abbot, 2001). These were labelled with 5'6-FAM™ fluorescent dye. PCRs were run in 20 μL reactions, consisting of 10 μL iTaq™, 1.5 μM primer and 1 μL DNA template (50–150 ng/μL). The PCR protocol was adapted from that of Saha et al. (2011) and Silva et al. (2013), and carried out as per Table S8. The presence of bands/peaks was validated by replicating each sample once in a different PCR machine (Taylor et al., 2011; Sutton et al., 2017). As a preliminary validation step, 5 μL PCR product was run on a 1.5% agarose gel at 6 V/cm for 3 h and viewed on a BioRad-Geldoc™ molecular imager system. DNA fragment analysis was performed using capillary electrophoresis (Applied Biosystems Inc., 3130 genetic analyser, GS500LIZ size standard) at the Central Analytical Facilities (CAF) division in Stellenbosch, South Africa.

2.7.2. ISSR data processing and analysis

Genemarker® v2.7.4 (SoftGenetics®) was used to read in and analyze all electropherograms, and generate binary data. RawGeno v2.0 (Arrigo et al., 2012) was subsequently used to further process this binary data such that the maximum and minimum bin widths were set to 1 and 0.5 base pairs (bp), respectively, scoring ranged from 100 to 500 bp, and a low relative fluorescent units (RFU) of 100 was applied, as recommended by AppliedBiosystems (2014) for data produced by 3130 Series instruments. Microsoft Excel® was used to organise the resulting binary matrix into replicate pairs, and prepare it for input into the BinMat v0.1.2 R package (van Steenderen, 2020) for replicate consolidation. This entailed that a '1' scored in one replicate, and a '0' scored at the same locus in the second replicate yielded a '?'. A '1' or a '0' in both replicates yielded a '1' and a '0', respectively.

SplitsTree4 v4.15.1 (Huson and Bryant, 2006) was used to generate NeighborNet trees from this consolidated data, applying the Jaccard's distance index. The NeighborNet method was chosen because Huson and Bryant (2006) suggested that it produces highly resolved networks. The Jaccard's distance index was applied because it does not consider the shared absence of bands as being biologically meaningful.

2.7.3. ISSR error rates

Two genotyping error rates were calculated, namely the 'Euclidean' (EE) and 'Jaccard' (JE) errors, as suggested by Bonin et al. (2004), Pompanon et al. (2005) and Holland et al. (2008). In the formulae below, f refers to the frequencies of the shared absence (00), shared presence (11), and ambiguities (01 and 10) of peaks in replicate sample pairs.

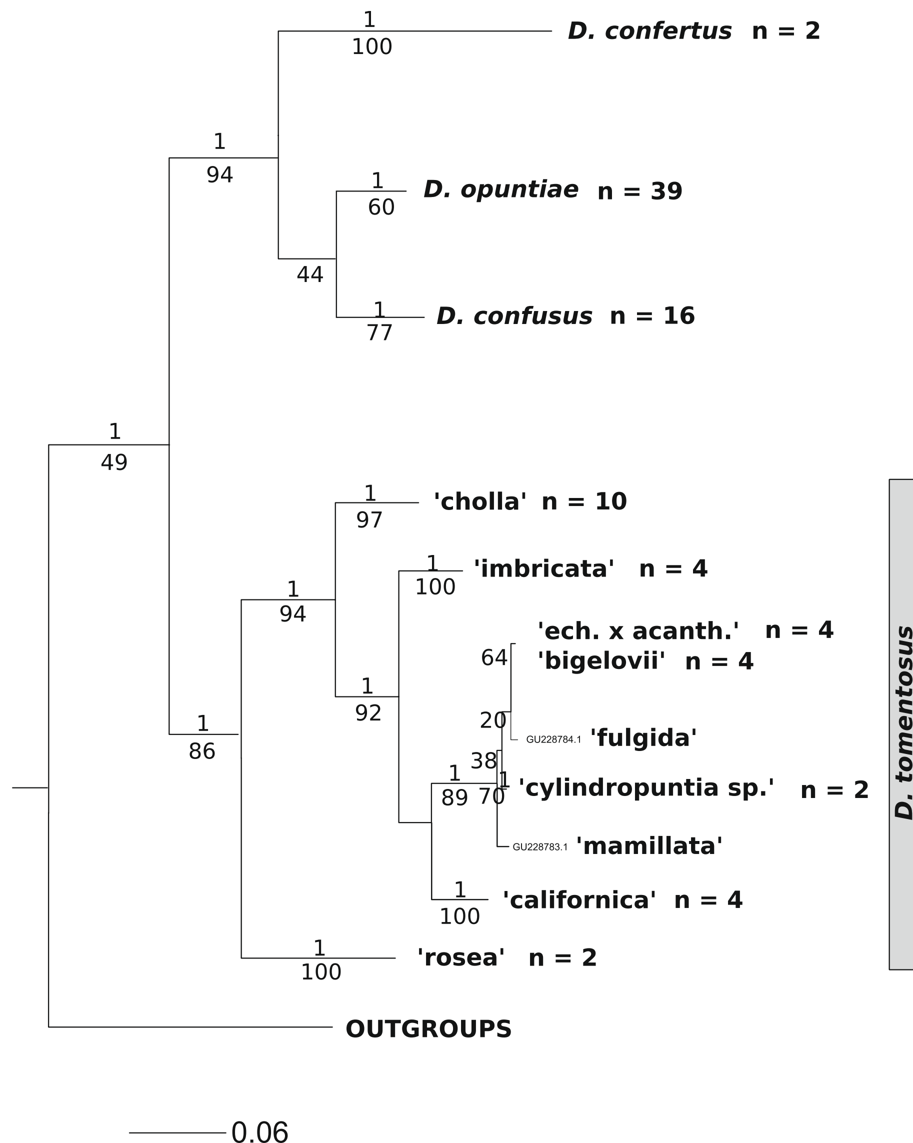


Fig. 1. Bayesian Inference (BI) and Maximum Likelihood (ML) phylogenies for the COI gene region, showing the BI topology. Posterior probabilities are above, and bootstrap values are below the branches. ML values are shown as a percentage of a thousand bootstrap replications. The scale bar (0.06) represents estimated substitutions per site.

$$EE = \frac{f_{10} + f_{01}}{f_{10} + f_{01} + f_{11} + f_{00}}$$

$$JE = \frac{f_{10} + f_{01}}{f_{10} + f_{01} + f_{11}}$$

2.8. Identification application via an R GUI interface

2.8.1. DactyID

The R Shiny library (RStudio, 2019) was used to create an online server enabling the identification of a query nucleotide sequence using the data collected in this study. The application ('DactyID') is available at https://clarkevansteenderen.shinyapps.io/Dactylopius_ID_version_1/ , or alternatively via GitHub on the user's local machine. The application can be accessed locally via GitHub by typing.

> runGitHub("DactyID", "CJMvS"). into the R console. A help file accompanies the program.

3. Results

3.1. Phylogenies

3.1.1. 12S

Sequence lengths were approximately 413 bp, with mean nucleotide base frequencies of A (44.57 ± 2.47%), C (12.79 ± 1.74%), T (38.11 ± 2.42%), and G (4.54 ± 2.08%). The Bayesian Inference (BI) and Maximum Likelihood (ML) phylogenies provided well supported clades for all the *Dactylopius* species included in the analysis, and for three of the six lineages within *D. tomentosus* (Fig. S1). The 'echinocarpa x acanthocarpa', 'bigelovii', and 'cylindropuntia sp.' lineages grouped together as one clade. There were only two nucleotide differences between 'bigelovii' and 'echinocarpa x acanthocarpa', and one nucleotide difference between the latter and 'cylindropuntia sp.'. The *D. opuntiae* 'ficus' and 'stricta' lineages did not show any separation (Fig. S1). The haplotype network for *D. opuntiae* shown in Fig. S1 illustrates this lack of unique haplotypic groups, with only seven segregating sites compared to 16 in the *D. confusus* network. Although with a range of only one to six nucleotide differences between haplotypes, the *D. confusus*

network indicates the existence of unique genetic groups across sites within the native range in the USA. Further investigation may reveal cryptic species within this complex. The *D. confertus* samples did not successfully amplify using this marker. The *D. ceylonicus* clade showed a well-supported split between the South African and Australian samples.

Apart from the outgroup, the greatest within-group K2P distances were *D. tomentosus* 'cholla' (0.068) and *D. ceylonicus* (0.05), followed by *D. confusus* (0.015) and *D. opuntiae* (0.0005) (Table S9). Excluding the outgroup, the average between-group K2P distance was 0.30 ± 0.14.

3.1.2. 18S

Sequence lengths were approximately 570 bp, with mean nucleotide base frequencies of A (25.02 ± 0.60%), C (24.17 ± 0.61%), T (22.95 ± 0.74%), and G (27.85% ± 0.40). The BI and ML phylogenies showed well-supported clades for all species, except for *Dactylopius austrinus* and *D. ceylonicus*, which grouped into one clade (Fig. S2). Within group K2P distances for the ingroups were all zero apart from the *D. austrinus* and *D. ceylonicus* group (0.006), and *D. tomentosus* (0.001) (Table S9). Excluding the outgroup, the average between-group K2P distance was 0.03 ± 0.02. This marker showed the *D. tomentosus* 'cholla' lineage forming a separate clade, and the 'echinocarpa x acanthocarpa' and 'bigelovii' grouping together, but the marker did not distinguish between the other three lineages (namely 'imbricata', 'californica var. parkeri', and 'cylindropuntia sp.'). It also did not show any intraspecific variation within the *D. opuntiae* clade.

3.1.3. COI

Sequence lengths were approximately 603 bp, with mean nucleotide base frequencies of A (36.33 ± 1.61%), C (19.64 ± 1.66%), T (37.55 ± 1.19%), and G (6.49 ± 0.55%).

Corroborating the 12S BI and ML phylogeny (Fig. 1 and S1), there is support for the 'echinocarpa x acanthocarpa', 'bigelovii', and 'cylindropuntia sp.' lineages forming one clade (Fig. 1 and S3). 'Bigelovii' samples differed from 'echinocarpa x acanthocarpa' by only one nucleotide (and thus did not separate into separate clades), and 'cylindropuntia sp.' samples differed from 'bigelovii' and 'echinocarpa x acanthocarpa' by 4, and 5 nucleotides, respectively (Fig. S3). As in the 12S haplotype network, *D. opuntiae* did not group into 'ficus' and 'stricta' haplotypes, and the *D. confusus* network again showed unique haplotypes for the different sites in the native range in the USA (with a larger range of nucleotide differences than 12S, ranging from 1 to 10 nucleotides) (Fig. S3).

The highest within-group K2P distance for the ingroup was for *D. confusus* (0.017), followed by *D. opuntiae* (0.014) and *D. confertus* (0.008) (Table S11). Excluding the outgroup, the average between-group K2P distance was 0.22 ± 0.1. The COI region also did not show enough genetic variation to separate the 'ficus' and 'stricta' lineages, or to distinguish between different intraspecific populations.

3.1.4. Concatenation of the 12S and 18S gene regions

The 12S and 18S data sets were concatenated due to a significant Icong value (Icong = 1.72, p = 2.47x10^{-8}), indicating that the gene tree topologies were significantly congruent. All the *Dactylopius* species, and half of the *D. tomentosus* lineages formed separate clades. The *D. tomentosus* 'bigelovii', 'cylindropuntia sp.', and 'echinocarpa x acanthocarapa' lineages grouped together, and the *D. opuntiae* 'ficus' and 'stricta' lineages did not separate (Figs. 2 and S4). Since none of the gene phylogenies (12S, 18S, COI, or concatenated 12S and 18S) could differentiate between the *D. opuntiae* 'ficus' and 'stricta' lineages, ISSR analyses were used to gain higher resolution results for this species.

3.2. Identification accuracy

3.2.1. 12S

At the species level, and at the optimal genetic distance threshold of 1%, identification accuracy (IA) was 100% for both barcode tests (BCM and TID) (Table 1). At the lineage level, *D. tomentosus* had an IA of 100% (BCM) and 82.4% (TID) at a distance threshold of 1%. The TID result increased to 100% at a lower optimal threshold value of 0.2%. *Dactylopius opuntiae* only had an IA of 15.22% (with ambiguities at 82.61% and incorrect IDs at 2.17%) for the BCM test and a zero IA (100% ambiguities) for the TID test at a threshold of 1%. At a decreased threshold of 0.2%, the TID only increased to a 13.04% IA, with ambiguities at 80.43% (incorrect and no IDs at 2.17% and 4.35%, respectively). The range of threshold genetic distance values for the *D. opuntiae* lineages all showed a high occurrence of false negatives. Barcode gaps at the species level and for *D. tomentosus* lineages were all positive (i.e., interspecific > intraspecific variation), while the lineages within *D. opuntiae* were all negative (i.e., interspecific < intraspecific variation). Of the *D. tomentosus* lineages, the 'cylindropuntia sp.' and 'echinocarpa x acanthocarpa' lineages had the smallest barcode gaps.

3.2.2. 18S

At the species level, and at the default genetic threshold of 1%, IA was 94.59% (5.41% ambiguity) for the BCM test, but only 29.73% (70.27% ambiguities) for the TID test (Table 1). At a lower threshold of 0.2%, the IA for both the BCM and TID tests were 94.59%, with ambiguities of 5.41%. At the lineage level, at a threshold of 0.1%, *D. tomentosus* had an IA of 22.22% (77.78% ambiguities) for both the BCM and TID tests. The BCM tests did not produce output for the intraspecific lineages of *D. opuntiae*, as they were indistinguishable. The TID tests for this lineage produced 100% ambiguous results at both the 1% and 0.1% thresholds. The range of genetic threshold distance values for the lineages within *D. tomentosus* and *D. opuntiae* all showed a high number of false negatives. Barcode gaps at the species level were all positive, except for *D. austrinus* and *D. ceylonicus* sequences, which had negative barcode gaps (as illustrated in the phylogenetic tree in Fig. S2). The largest positive barcode gaps at the species level were for *D. tomentosus*. Barcode gaps for *D. opuntiae* and *D. tomentosus* at the lineage level were all zero (i.e., the intra-and interspecific variation in these groups were equal); except for *D. tomentosus* 'cholla' sequences, which had positive barcode gaps.

3.2.3. COI

Identification accuracy at the species level at a 1% threshold was 100% for both the BCM and TID tests (Table 1). At the lineage level, at a 1% threshold, *D. opuntiae* had an IA of 61.54% (32.69% ambiguities and 5.77% incorrect) for the BCM test, and an IA of 46.15% (51.92% ambiguities and 1.92% incorrect) for the TID test. The BCM test results remained the same at a lower optimal threshold of 0.8%, but the TID IA rose to 59.62% (38.46% ambiguities and 1.92% incorrect). At the lineage level for *D. tomentosus*, when 'bigelovii', 'cylindropuntia sp.', and 'echinocarpa x acanthocarpa' were treated as separate lineages, IA was 96.3% (3.7% ambiguities) for the BCM test, and 59.26% (37.04% ambiguities and 3.7% no ID) for the TID test (Table 1). At a lower threshold value of 0.2%, the BCM decreased to an IA of 88.89% (11.11% ambiguities), and the TID rose to 70.37% (18.52% ambiguities and 11.11% no ID). When 'bigelovii', 'cylindropuntia sp.', and 'echinocarpa x acanthocarpa' were treated as one group, at a threshold of 3.3%, the IA for both the BCM and TID tests was 100%. Barcode gaps were positive for all sequences at the species level, but negative for all *D. opuntiae* sequences at the lineage level. Barcode gaps for *D. tomentosus* lineages were all positive except for three 'bigelovii' sequences.

3.3. ISSR fragment analysis

Dactylopius opuntiae 'stricta' and 'ficus' lineages formed separate groups (Fig. 3). ISSR primer statistics are shown in Table 2. At the species level, at an optimal threshold of 60%, all barcode tests (BCM and TID) had a 100% IA (Table 3). At the lineage level, at an optimal threshold of 45%, *D. tomentosus* had an IA of 100% for both the BCM

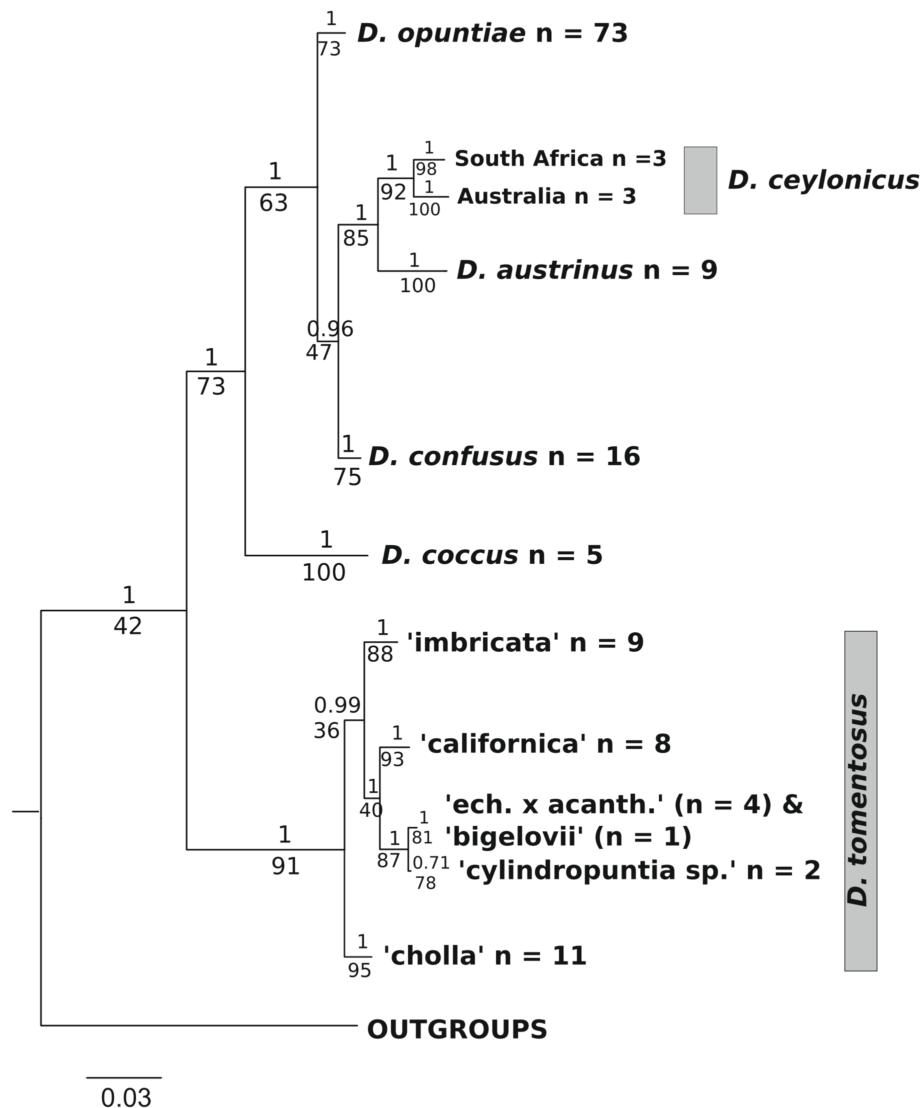


Fig. 2. Concatenated gene phylogenies for the 12S and 18S gene regions using Bayesian Inference (BI) and Maximum Likelihood (ML) methods, showing the BI topology. Bayesian posterior probability values are above, and ML bootstrap values are below the branches. ML values are shown as a percentage of a thousand bootstrap replications. The scale bar (0.03) represents estimated substitutions per site.

and TID tests (Table 3). At the same 45% threshold, *D. opuntiae* had an IA of 81.82% (no ID of 18.18%) for both the BCM and TID tests.

4. Discussion

4.1. DNA barcoding

For DNA barcoding to be useful it should be repeatable, relatively inexpensive, quick, and accurate. In this study, we sequenced three different gene regions and used two ISSRs to identify the different taxa and lineages of cochineal used as biocontrol agents. It is not necessary to use all of these regions and techniques to identify the different agent species or lineages used for biological control, and so in most cases, only one region or ISSR could be used to accurately identify the agent.

The 12S marker is recommended as the most efficient for *Dactylopius* species-level identification, and for the identification of the 'cholla', 'californica var. parkeri', and 'imbricata' lineages within the *D. tomentosus* species. The 12S rRNA primers successfully amplified the target region of five of the six *Dactylopius* species used in this project (only *D. confertus* was unsuccessful), and had the highest mean between-group K2P distance. The marker is not, however, useful for distinguishing between *D. opuntiae* 'ficus' and 'stricta' lineages, due to the presence of negative barcode gaps (i.e. intraspecific variation > interspecific variation) at that higher-resolution taxonomic level. The 12S marker, in combination with a variety of other informative mitochondrial genes (such as 16S rRNA and cytochrome B), may also help in future to determine the dates at which lineages of cochineals have been separated by means of molecular clock dating. The two *D. ceylonicus* lineages from Australia and South Africa, for example, showed a well-supported separation. These lineages are from the same stock, and they are known to have been isolated for 105 years (Winston et al., 2014), which could be used to estimate how long other cochineal lineages have been isolated. The 18S rRNA marker could accurately distinguish between all *Dactylopius* species, except for *D. austrinus* and *D. ceylonicus*, which grouped into one clade and were the only two species with negative barcode gaps. Of the *D. tomentosus* lineages, only 'cholla', and 'echinocarpa x acanthocarpa' and 'bigelovii' formed separate clades, while 'imbricata', 'californica var. parkeri', and 'cylindropuntia sp.' were unresolved. In the study by Campana et al., 2015, no intraspecific variation was found within the seven 18S sequences obtained from *D. coccus* populations. The findings of the current project corroborate with those of Campana et al., 2015, namely that this marker is uninformative for identification beyond the species level. It can, however, be used for the inference of phylogenetic relationships between species,

Table 1
Results of the Best Close Match (BCM) and Threshold ID (TID) barcode testing algorithms for the 12S, 18S, and COI gene regions. Values for BCM and TID are shown at the default 1% and at the optimum threshold value. Results are shown at the species and lineage level. In the COI section, *D. tomentosus* (G) indicates a test conducted where the 'bigelovii', 'cylindropuntia sp.', and 'echinocarpa x acanthocarpa' sequences were grouped as one lineage.

		BCM (1%)	Proportion of Samples (%)	BCM (optimum %)	Proportion of Samples (%)	TID (1%)	Proportion of Samples (%)	TID (optimum %)	Proportion of Samples (%)
12S				**1% threshold**				**1% threshold**	
Species level	**Correct**	147	100	147	100	147	100	147	100
Lineage level	**D. opuntiae**			**0.2% threshold**				**0.2% threshold**	
	Correct	7	15.22	6	13.04	0	0	6	13.04
	Incorrect	1	2.17	1	2.17	0	0	1	2.17
	Ambiguous	38	82.61	37	80.43	46	100	37	80.43
	No ID	0	0	2	4.35	0	0	2	4.35
	D. tomentosus			**1% threshold**				**0.2% threshold**	
	Correct	34	100	34	100	28	82.4	34	100
	Incorrect	0	0	0	0	0	0	0	0
	Ambiguous	0	0	0	0	6	17.6	0	0
18S				**0.2% threshold**				**0.2% threshold**	
Species level	**Correct**	140	94.59	140	94.59	44	29.73	140	94.59
	Ambiguous	8	5.41	8	5.41	104	70.27	8	5.41
Lineage level	**D. opuntiae**							**0.1% threshold**	
	Correct	BCM test error due to identical sequences shared across test groups.				0	0	0	0
	Incorrect					0	0	0	0
	Ambiguous					35	100	35	100
	No ID					0	0	0	0
	D. tomentosus			**0.1% threshold**				**0.1% threshold**	
	Correct	8	22.22	8	22.22	0	0	8	22.22
	Incorrect	0	0	0	0	0	0	0	0
	Ambiguous	28	77.78	28	77.78	36	100	28	77.78
COI				**1% threshold**				**1% threshold**	
Species level	**Correct**	83	100	83	100	83	100	83	100
Lineage level	**D. opuntiae**			**0.8% threshold**				**0.8% threshold**	
	Correct	32	61.54	32	64.54	24	46.15	31	59.62
	Incorrect	3	5.77	3	5.77	1	1.92	1	1.92
	Ambiguous	17	32.69	17	32.69	27	51.92	20	38.46
	D. tomentosus			**0.2% threshold**				**0.2% threshold**	
	Correct	26	96.3	24	88.89	16	59.26	19	70.37
	Incorrect	0	0	0	0	0	0	0	0
	Ambiguous	0	0	0	0	10	37.04	5	18.52
	No ID	1	3.7	3	11.11	1	3.7	3	11.11
	D. tomentosus (G)			**3.3% threshold**				**3.3% threshold**	
	Correct	26	96.3	27	100	26	96.3	27	100
	Incorrect	0	0	0	0	0	0	0	0
	Ambiguous	0	0	0	0	0	0	0	0
	No ID	1	3.7	0	0	1	3.7	0	0

particularly when supplementing a concatenated genetic data set.

COI-A primers could distinguish between *D. confertus*, *D. opuntiae*, and *D. confusus* with a 100% IA. A 1% genetic distance threshold was found to be to be optimal for assessing identification accuracy at the species level, which is consistent with the barcoding literature (Hebert et al., 2003; Hebert et al., 2003). It is not, however, able to distinguish between *D. opuntiae* lineages. The 'rosea' lineage appeared as an outgroup to the *D. tomentosus* clade, which might relate to its failure to establish on any *Cylindropuntia* host plants in Australia (Jones et al., 2015).

The COI results corroborate the 12S and 18S phylogenies, with the *D. tomentosus* 'echinocarpa x acanthocarpa', 'bigelovii', and 'cylindropuntia sp.' lineages clustered tightly together. This could be due to geography, as these three lineages were collected in Arizona in the United States of America between 2012 and 2015 (Jones et al., 2016a). The 'rosea', 'cholla', and 'californica var. parkeri' lineages were collected in Mexico, and the 'imbricata' lineage was originally collected in Texas, USA (Winston et al., 2014).

The *D. opuntiae* ISSR data presented here showed clear differences between the 'ficus' and 'stricta' lineages that the traditional DNA gene regions failed to reveal. Due to the different kind of data that ISSR analyses produce, the distance threshold required to create taxonomic group designations was much higher than that of nucleotide sequence data (60% compared to the standard 1%). This is because the latter is much more data-rich compared to binary information, such as that produced by ISSR analyses.

ISSR analysis using the methods described in this work offers a new and valuable tool to biological control efforts, as it allows practitioners to distinguish between these otherwise morphologically identical lineages. The *D. tomentosus* lineages included in the ISSR analyses

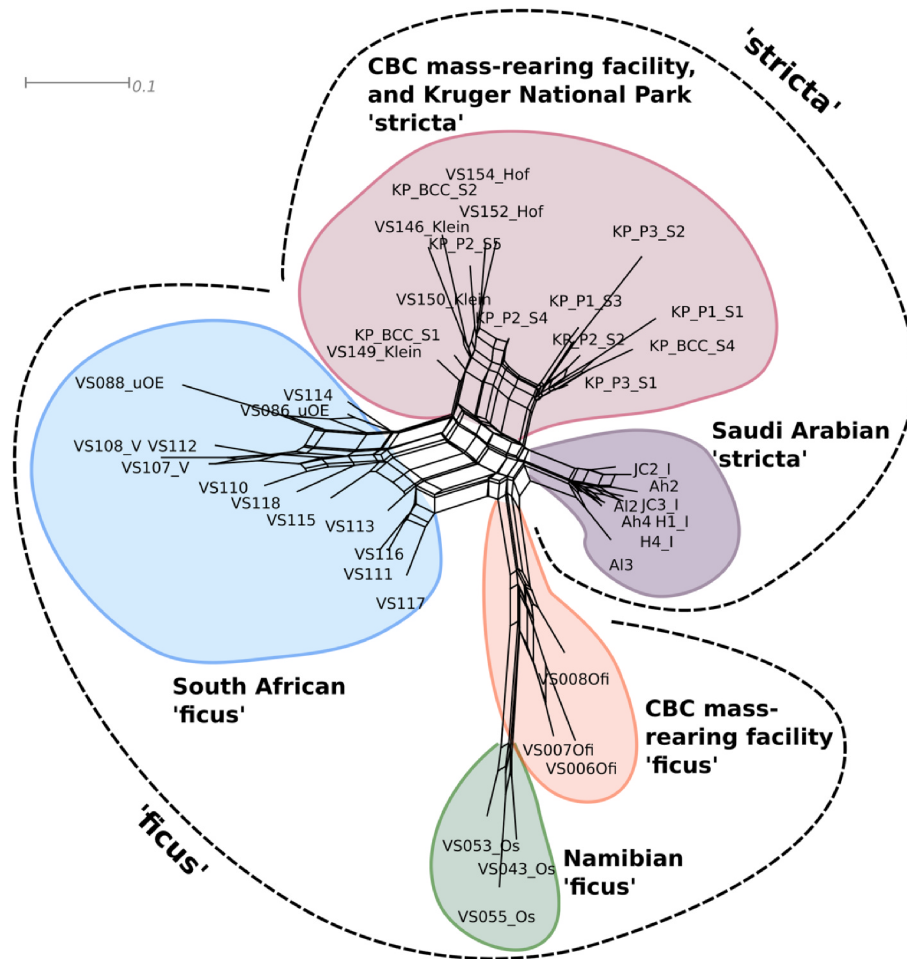


Fig. 3. SplitsTree graphical output (NeighborNet method, applying Jaccard's index to calculate the distance matrix) for *Dactylopius opuntiae* individuals using two ISSRs (ISSR 809 and ISSR 826). The scale bar (0.1) represents the number of substitutions per site.

('imbricata', 'cholla', and 'californica var. parkeri') showed a 100% IA at a distance threshold of 45%, although using the 12S rRNA marker to identify these is less labour intensive. Although a 82% hit rate is high for the identification of *D. opuntiae* lineages, there are a multitude of other ISSR primers that may offer lower error rates and amplify a larger number of fragments. It would be worth using capillary electrophoresis to test the TCT(GA)$_7$ ISSR primer reported by Silva et al., 2013 as producing the highest percentage of polymorphic loci, and to test the RAPD primers used in their study.

Of the two ISSR primers used here, ISSR 826 had an average Jaccard error rate 5% greater than ISSR 809, and tended to be less informative for some population groups compared to the latter primer. ISSR 809 is therefore recommended as the better of the two primers for identification of the *D. opuntiae* lineages. On the whole, ISSR analyses are more time-consuming and labor-intensive than traditional DNA barcoding methods, so for all lineages except the two *D. opuntiae* lineages, the use of either the 12S or COI region would be preferable.

4.2. Practical applications for cactus biological control using cochineal insects

The barcoding methods used in this study provide a wide range of practical applications for the identification of the Dactylopiidae; particularly for countries such as South Africa and Australia that are the most reliant on biological control to reduce cactus invasions. The identification techniques provided here make it possible to identify field-collected specimens, and to plan releases of the most effective agent onto the correct target weed. It could also help to explain differences in agent performance; agents might have been moved, or even dispersed onto, a less preferred host. Additionally, these techniques could be used to confirm whether or not a newly-released lineage has established in the field, where it may co-occur with a previously-released lineage that is not causing significant impact.

Intra-specific lineages within the Dactylopiidae are known to hybridise. This directly affects the host specificity of hybrid populations,

Table 2
Summary of the number of loci, the average number of bands obtained, standard deviations, the maximum and minimum number of peaks yielded, and the Euclidean and Jaccard error rates for ISSR primers 809 and 826. Values were obtained after conservative filtering parameters had been applied to the data in RawGeno, and are representative of all the individuals for each primer's data set.

Primer	No. loci	Avg. peaks ± sd	Max. peaks	Min. peaks	Avg. Euclidean error rate ± sd (Bonin et al., 2004)	Avg. Jaccard error rate ± sd (Holland et al., 2008)
ISSR 809	384	36.58 ± 15.75	88	5	0.08 ± 0.03	0.48 ± 0.14
ISSR 826	366	32.09 ± 18.93	87	4	0.09 ± 0.03	0.53 ± 0.17

Table 3
Barcode testing results for the Best Close Match (BCM) and Threshold ID (TID) tests, at the species and lineage level.

		BCM	% of samples	TID	% of samples
Species level		**60% threshold**		**60% threshold**	
	Correct	124	100	124	100
	No ID	0	0	0	0
Lineage level	**D. tomentosus**	**45% threshold**		**45% threshold**	
	Correct	10	100	10	100
	Incorrect	0	0	0	0
	Ambiguous	0	0	0	0
	D. opuntiae	**45% threshold**		**45% threshold**	
	Correct	36	81.82	36	81.82
	Incorrect	0	0	0	0
	Ambiguous	0	0	0	0
	No ID	8	18.18	8	18.18

and possibly their virulence. The work done by Hoffmann et al., 2002 and Mathenge et al., 2010 on lineages of *D. opuntiae* and *D. tomentosus*, respectively, are good examples of this. Hybrid performance has important implications in the planning of agent releases; as lineages that produce less damaging hybrid offspring should be kept apart in the field and restricted to the correct target weed species. The dynamics of the hybridisation between other *Dactylopius* lineages are still unknown. Fragment analysis methods, such as ISSRs, could be a useful identification tool in such future studies.

Countries such as South Africa and Australia are eager to source and test new agents for controlling target Cactaceae, and these techniques could help in the identification of new species, cryptic species, and lineages used in future biological control programmes. Both Paterson et al., 2011 and Jones et al., 2016a highlight the need for multiple *Dactylopius* lineages to control a range of species. The Dactylopiidae are notoriously difficult to identify, and taxonomists can sometimes misidentify specimens due to problems with this family's taxonomy and identification keys. DNA and molecular data offers an independent line of evidence for identification, and also provides phylogenetic information about the relationships between species and lineages, and their geographical distributions.

Dactylopius opuntiae is considered a useful biological control agent, but it can also be considered a pest on *O. ficus-indica* crops. Therefore, it is important to be able to distinguish between the *D. opuntiae* 'stricta' and 'ficus' cochineal lineages, because although they are the same species, the 'stricta' cochineal is only a biocontrol agent and will not feed on *O. ficus-indica* (Githure et al., 1999), while the 'ficus' lineage could be a biocontrol agent for several *Opuntia* species or a pest, because it feeds on *O. ficus-indica*. In other words, the 'ficus' cochineal lineage could be considered an effective biological control agent in areas where cacti are problematic, and an agricultural pest where they are grown as crops.

Having genetic tools to distinguish between what is a potential pest, and what is a beneficial biological control agent could be very useful. Especially since the reputation of the biological control of weeds as a safe practice is at stake. With the exception of the 'ficus' lineage, all cochineal species and lineages used for biological control will not feed on the *O. ficus-indica* crop, but without being able to identify the different species and lineages of cochineal, false claims of unpredicted non-target impacts from cochineals used as biological control agents could arise.

5. Conclusion

The taxonomic history of the Dactylopiidae is riddled with misidentifications, and so a genetic approach is a much-needed addition to the identification toolkit of the Dactylopiidae. This can assist in detecting new species, cryptic species, and lineages. The control of invasive Cactaceae is one of the most successful biological control initiatives in South Africa and Australia, and stands to gain substantially from the streamlined and accurate identification process presented here. Most importantly, this can ensure that the correct and most damaging cochineal is used on the correct target weed, thus maximising the level of control that is obtained.

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.

CRediT authorship contribution statement

C.J.M. van Steenderen: Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Visualization. **I.D. Paterson:** Conceptualization, Resources, Writing - review & editing, Supervision, Funding acquisition. **S. Edwards:** Resources, Writing - review & editing, Supervision. **M.D. Day:** Writing - review & editing, Funding acquisition.

Acknowledgments

Funding was provided by the Working for Water (WfW) programme of the Department of Environment, Forestry and Fisheries: Natural Resource Management programme (DEFF: NRM). Funding for this work was also provided by the South African Research Chairs Initiative of the Department of Science and Technology, the National Research Foundation (NRF) of South Africa, and the Queensland Department of Agriculture and Fisheries, Australia. Any opinion, finding, conclusion or recommendation expressed in this material is that of the authors and the NRF does not accept any liability in this regard.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, athttps://doi.org/10.1016/j.biocontrol.2020.104426.

References

Abbot, P., 2001. Individual and population variation in invertebrates revealed by Inter-simple Sequence Repeats (ISSRs). J. Insect Sci. 1, 8.
AppliedBiosystems. DNA fragment analysis by capillary electrophoresis. Applied Biosystems, 2014.
Arrigo, N., Holderegger, R., Alvarez, N., 2012. Automated scoring of AFLPs using RawGeno v 2.0, a free R CRAN library. In: Data Production and Analysis in Population Genomics, Springer, pp. 155–175. doi: 10.1007/978-1-61779-870-2_10.
Birch, J.L., Walsh, N.G., Cantrill, D.J., Holmes, G.D., Murphy, D.J., 2017. Testing efficacy of distance and tree-based methods for DNA barcoding of grasses (Poaceae tribe Poeae) in Australia. Public Lib. Sci. One 12 (10), e0186259. https://doi.org/10.1371/journal.pone.0186259.
Bonin, A., Bellemain, E., Bronken Eidesen, P., Pompanon, F., Brochmann, C., Taberlet, P., 2004. How to track and assess genotyping errors in population genetics studies. Mol. Ecol. 13 (11), 3261–3273. https://doi.org/10.1111/j.1365-294X.2004.02346.x.
Brown, C.J., Macdonald, I.A.W., Brown, S., 1985. Invasive alien organisms in South West Africa/Namibia. Foundation for Research Development: CSIR.
Brown, S.D.J., Collins, R.A., Boyer, S., Lefort, M.-C., Malumbres-Olarte, J., Vink, C.J., Cruickshank, Robert H., 2012. Spider: an R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. Mol. Ecol. Resour. 12 (3), 562–565. https://doi.org/10.1111/j.1755-0998.2011.03108.x.
Bruford, M.W., Hanotte, O., Brookfield, J.F.Y., Burke, T., 1992. Single locus and multi-locus DNA fingerprinting. In: Hoezel, C.A.R. (Ed.), Molecular Genetics Analysis of Populations: A Practical Approach. Oxford University Press, New York, pp. 225–269.
Campana, M.G., Robles García, N.M., Tuross, N., 2015. America's red gold: multiple lineages of cultivated Cochineal in Mexico. Ecol. Evol. 5 (3), 607–617. https://doi.org/10.1002/ece3.1398.
Clement, M., Posada, D., Crandall, K.A., 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9 (10), 1657–1659. https://doi.org/10.1046/j.1365-294x.

2000.01020.x.
Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9 (8). https://doi.org/10.1038/nmeth.2109. 772–772.
De Lotto, G., 1974. On the status and identity of the cochineal insects (Homoptera: Coccoidea: Dactylopiidae). J. Entomol. Soc. Southern Afr. 37 (1), 167–193.
de Vienne, D.M., Giraud, T., Martin, O.C., 2007. A congruence index for testing topological similarity between trees. Bioinformatics 23 (23), 3119–3124. https://doi.org/10.1093/bioinformatics/btm500.
Downie, D.A., 2010. Baubles, bangles, and biotypes: a critical review of the use and abuse of the biotype concept. J. Insect Sci. 10 (176), 1–18. https://doi.org/10.1673/031.010.14136.
Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17 (6), 368–376. https://doi.org/10.1007/BF01734359.
Felsenstein, J. Consense (v3.69), 2006. http://www.phylo.org/index.php/rest/consense.html.
Githure, C., Zimmermann, H., Hoffmann, J., 1999. Host specificity of biotypes of *Dactylopius opuntiae* (Cockerell) (Hemiptera: Dactylopiidae): prospects for biological control of *Opuntia stricta* (Haworth) (Cactaceae) in Africa. Afr. Entomol. 7 (1), 43–48.
Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52 (5), 696–704. https://doi.org/10.1080/10635150390235520.
Gullan, P.J., Kosztarab, M., 1997. Adaptations in Scale Insects. Annu. Rev. Entomol. 42 (1), 23–50. https://doi.org/10.1146/annurev.ento.42.1.23.
Hall, T., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In: Nucleic Acids Symposium Series. Oxford University Press, Raleigh, North Carolina, USA, pp. 95–98.
Hebert, P.D.N., Cywinska, A., Ball, S.L., deWaard, J.R., 2003. Biological Identifications through DNA Barcodes. Proc. Roy. Soc. B Biol. Sci. 270 (1512), 313–321. https://doi.org/10.1098/rspb.2002.2218.
Hebert, P.D.N., Ratnasingham, S., de Waard, J.R., 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc. Roy. Soc. B Biol. Sci. 270 (Suppl_1), S96–S99. https://doi.org/10.1098/rsbl.2003.0025.
Hoffmann, J.H., Impson, F.A.C., Volchansky, C.R., 2002. Biological control of Cactus weeds: implications of hybridization between control agent biotypes: performance of hybrid cochineal biotypes. J. Appl. Ecol. 39 (6), 900–908. https://doi.org/10.1046/j.1365-2664.2002.00766.x.
Holland, B.R., Clarke, A.C., Meudt, H.M., 2008. Optimizing automated AFLP scoring parameters to improve phylogenetic resolution. Syst. Biol. 57 (3), 347–366. https://doi.org/10.1080/10635150802044037.
Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17 (8), 754–755. https://doi.org/10.1093/bioinformatics/17.8.754.
Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23 (2), 254–267. https://doi.org/10.1093/molbev/msj030.
Jones, P.K., Holtkamp, R.H., Palmer, W.A., Day, M.D., 2015. The host range of three biotypes of *Dactylopius tomentosus* (Lamarck) (Hemiptera: Dactylopiidae) and their potential as biological control agents of *Cylindropuntia* Spp. (Cactaceae) in Australia. Biocontrol Sci. Technol. 25 (6), 613–628. https://doi.org/10.1080/09583157.2014.999747.
Jones, P., Holtkamp, R., Day, M., 2016a. The host range of four new biotypes of *Dactylopius tomentosus* (Hemiptera: Dactylopiidae) from southern USA and their potential as biological control agents of *Cylindropuntia* spp. (Cactaceae) in Australia: Part II. Biocontrol Sci. Technol. 26 (8), 1033–1047. https://doi.org/10.1080/09583157.2016.1182620.
Jones, P., Oakey, J., Day, M., Vitelli, J., 2016b. Improving the effectiveness of the release strategy for the array of *Dactylopius tomentosus* biotypes for the biocontrol of *Cylindropuntia* species by using DNA analysis. In: Randall, R., Lloyd, S., and Borger, C. (Eds.), 20th Australasian Weeds Conference 2016. Weed Society of Western Australia, 2016b. ISBN 9781510834088.
Kaplan, H., Wilson, J.R., Klein, H., Henderson, L., Zimmermann, H.G., Manyama, P., Ivey, P., Richardson, D.M., Novoa, A., 2017. A proposed national strategic framework for the management of Cactaceae in South Africa. Bothalia – African Biodivers. Conserv. 47 (2), 1–12. https://doi.org/10.4102/abc.v47i2.2149.
Katoh, K., Rozewicki, J., Yamada, K.D., 2017. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings Bioinf. https://doi.org/10.1093/bib/bbx108.
Klein, H., 2011. A catalogue of the insects, mites and pathogens that have been used or rejected, or are under consideration, for the biological control of invasive Alien Plants in South Africa. Afr. Entomol. 19 (2), 515–549. https://doi.org/10.4001/003.019.0214.
Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis Version 7.0 for bigger datasets. Mol. Biol. Evol. 33 (7), 1870–1874. https://doi.org/10.1093/molbev/msw054.
Leigh, J.W., Bryant, D., 2015. POPART: full-feature software for haplotype network construction. Methods Ecol. Evol. 6 (9), 1110–1116. https://doi.org/10.1111/2041-210X.12410.
Mann, J., 1969. Cactus-feeding insects and mites. Bull. United States Nat. Museum 256 (256), 1–158. https://doi.org/10.5479/si.03629236.256.1.
Mathenge, C.W., Holford, P., Hoffmann, J.H., Zimmermann, H.G., Spooner-Hart, R., Beattie, G.A.C., 2010. Determination of biotypes of *Dactylopius tomentosus* (Hemiptera: Dactylopiidae) and insights into the taxonomic relationships of their hosts *Cylindropuntia* spp. Bull. Entomol. Res. 100 (03), 347–358. https://doi.org/10.1017/S0007485309990496.
Mathenge, C.W., Riegler, M., Beattie, G.A.C., Spooner-Hart, R.N., Holford, P., 2015. Genetic variation amongst biotypes of *Dactylopius tomentosus*. Insect Sci. 22 (3), 360–374. https://doi.org/10.1111/1744-7917.12120.
Meier, R., Shiyang, K., Vaidya, G., Ng, P.K.L., 2006. DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. Syst. Biol. 55 (5), 715–728. https://doi.org/10.1080/10635150600969864.
Miller, M., 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), New Orleans, LA, pp. 1–8.
Novoa, A., Le Roux, J.J., Robertson, M.P., Wilson, J.R.U., Richardson, D.M., 2015. Introduced and invasive Cactus species: a global review. Ann. Botany Plants 7 (0): plu078–plu078. doi: 10.1093/aobpla/plu078.
Paradis, E., Claude, J., Strimmer, K., 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20 (2), 289–290. https://doi.org/10.1093/bioinformatics/btg412.
Park, D.-S., Suh, S.-J., Oh, H.-W., Hebert, P.D.N., 2010. Recovery of the mitochondrial COI Barcode region in diverse Hexapoda through tRNA-based primers. BioMed Central Genom. 11, 7. https://doi.org/10.1186/1471-2164-11-423.
Paterson, I.D., Hoffmann, J.H., Klein, H., Mathenge, C.W., Neser, S., Zimmermann, H.G., 2011. Biological control of Cactaceae in South Africa. Afr. Entomol. 19 (SP), 230–246. https://doi.org/10.4001/003.019.0221.
Pérez-Guerra, G., Kosztarab, M., 1992. Biosystematics of the family Dactylopiidae (Homoptera: Coccinea) with emphasis on the life cycle of *Dactylopius coccus* Costa. PhD thesis, Virginia Polytechnic Institute and State University.
Pompanon, F., Bonin, A., Bellemain, E., Taberlet, P., 2005. Genotyping errors: causes, consequences and solutions. Nat. Rev. Genet. 6 (11), 847. https://doi.org/10.1038/nrg1707.
Portillo, M., Vigueras, A., 2006. A Review on the cochineal species in Mexico, hosts and natural enemies. Acta Hortic. 728 (728), 249–256. https://doi.org/10.17660/ActaHortic.2006.728.35.
R Core Team. R: a language and environment for statistical computing, 2019. URL http://www.R-project.org/.
Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67 (5), 901–904. https://doi.org/10.1093/sysbio/syy032.
Ramírez-Puebla, S.T., Rosenblueth, M., Chávez-Moreno, C.K., Pereira, Catanho, de Lyra, M.C., Tecante, A., Martínez-Romero, E., 2010. Molecular phylogeny of the genus *Dactylopius* (Hemiptera: Dactylopiidae) and identification of the symbiotic bacteria. Environ. Entomol. 39 (4), 1178–1183. https://doi.org/10.1603/EN10037.
RStudio, Inc., Easy web applications in R., 2019. URL: http://www.rstudio.com/shiny/.
RStudio Team. RStudio: Integrated Development Environment for R. RStudio, Inc., Boston, MA, 2015.
Saha, D., Ranjan, S.K., Mallick, C.B., Vidyarthi, A.S., Ramani, R., 2011. Genetic diversity in lac resin-secreting insects belonging to *Kerria* spp., as revealed through ISSR markers. Biochem. Syst. Ecol. 39 (2), 112–120. https://doi.org/10.1016/j.bse.2011.01.016.
Sheehan, M., Potter, S., 2017. Managing Opuntioid Cacti in Australia: best practice control manual for Austrocylindropuntia, Cylindropuntia and Opuntia species. Department of Primary Industries and Regional Development 6, 0, 0. WA, Perth.
Silva, D., Mergulhão, C., Medeiros, L., Figueiredo, M., Burity, H., 2013. Genetic variability of *Dactylopius opuntiae* (Hemiptera, Dactylopiidae) on forage cactus in northeast Brazil. Genet. Mol. Res. 12 (4), 5236–5246. https://doi.org/10.4238/2013.October.30.8.
Sutton, G.F., Paterson, I.D., Paynter, Q., 2017. Genetic matching of invasive populations of the African Tulip tree, *Spathodea campanulata* Beauv. (Bignoniaceae), to their native distribution: maximising the likelihood of selecting host-compatible biological control agents. Biol. Control 114, 167–175. https://doi.org/10.1016/j.biocontrol.2017.08.015.
Taylor, S.J., Downie, D.A., Paterson, I.D., 2011. Genetic diversity of introduced populations of the water Hyacinth biological control agent *Eccritotarsus catarinensis* (Hemiptera: Miridae). Biol. Control 58 (3), 330–336. https://doi.org/10.1016/j.biocontrol.2011.05.008.
Van Dam, A.R., Portillo Martinez, L., Chavez, A.J., May, B.P., 2015. Range wide phylogeography of *Dactylopius coccus* (Hemiptera: Dactylopiidae). Ann. Entomol. Soc. Am. 108 (3), 299–310. https://doi.org/10.1093/aesa/sav017.
van Steenderen, C.J.M., 2020. BinMat: Processes Binary Data Obtained from Fragment Analysis. URL https://cran.r-project.org/web/packages/BinMat/. R package version 0.1.2.
Volchansky, C.R., Hoffmann, J.H., Zimmermann, H.G., 1999. Host-plant affinities of two biotypes of *Dactylopius opuntiae* (Homoptera: Dactylopiidae): enhanced prospects for biological control of *Opuntia stricta* (Cactaceae) in South Africa. J. Appl. Ecol. 36 (1), 85–91. https://doi.org/10.1046/j.1365-2664.1999.00381.x.
Winston, R.L., Schwarzländer, M., Hinz, H.L., Day, M.D., Cock, M.J.W., Julien, M.H., 2014. Biological control of weeds: a world catalogue of agents and their target weeds., vol. 5. USDA Forest Service, Forest Health Technology Enterprise Team.
Xia, X., 2018. DAMBE7: new and improved tools for data analysis in molecular biology and evolution. Mol. Biol. Evol. 35 (6), 1550–1552. https://doi.org/10.1093/molbev/msy073.
Xia, X., Lemey, P., 2009. Assessing substitution saturation with DAMBE. In: Lemey, P., Salemi, M., Vandamme, A.M. (Eds.), The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny. Cambridge University Press, pp. 615–630.
Xia, X., Xie, Z., Salemi, M., Chen, L., Wang, Y., 2003. An index of substitution saturation and its application. Mol. Phylogenet. Evol. 26 (1), 1–7. https://doi.org/10.1016/S1055-7903(02)00326-3.
Zwickl, D.J., 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. PhD thesis, University of Texas at Austin.