



**RHODES UNIVERSITY**  
*Where leaders learn*

---

**A genetic analysis of the species  
and intraspecific lineages of *Dactylopius*  
Costa (Hemiptera: Dactylopiidae)**

---

By

**Clarke van Steenderen**

A thesis submitted in fulfilment of the requirements for the degree of  
Masters in Entomology

Department of Zoology and Entomology  
Centre for Biological Control

December 2019

This thesis is dedicated to my loving mother,  
Thea,  
who has always encouraged and loved me unconditionally

# Declaration

I hereby declare that this thesis has not been submitted, either in the same or different form, to this or any other university for a degree and that it represents my own work. I know the meaning of plagiarism and declare that all of the work in the thesis, save for that which is properly acknowledged, is my own. The ethics clearance number for this project is 2019-0358-242.

---

Clarke van Steenderen

December 2019

# Acknowledgements

I would like to express my gratitude to the following people and organisations:

The National Research Foundation (NRF) for funding the bulk of my M.Sc. degree.

Rhodes University and the Levenstein family for the provision of bursaries.

A special note of thanks to Mr. John Gillam at the Rhodes University Postgraduate funding office for all his help with organising funding.

My supervisor, Dr. Iain Paterson, for his valuable feedback and encouragement. Iain's enthusiasm is contagious!

My co-supervisor, Dr. Shelley Edwards, for her valuable advice, and assistance in the Zoology and Entomology Molecular Lab (ZEMML) at Rhodes University.

Prof. Martin Hill for all his support, positivity, and encouragement. The Entomology Department at Rhodes University is incredibly lucky to have someone as resourceful and passionate as he is.

The students and staff at the Centre for Biological Control (CBC) and the Department of Zoology and Entomology at Rhodes, with a special thanks to Mrs. Jeanne van der Merwe for her assistance in organising payments and other administrative work.

The Central Analytical Facilities (CAF) at Stellenbosch University for the fragment analysis of ISSR samples, with a special thanks to Mrs. René Veikondis.

My office mates: Ant, Evans, and Megan, for all the laughs and fun we have had.

My flat mates: Elisa, Glyn, and Olivier, for their friendship.

Lastly, my mother, and the family and friends who have shown their love and support.

# Abbreviations, Symbols, and Units

## Abbreviations

AFLP	Amplified Fragment Length Polymorphism
ANOSIM	Analysis of Similarity
ARC	Agricultural Research Council
BCM	Best Close Match
BI	Bayesian Inference
BOLD	Barcode of Life Database
bp	Base Pairs
BPP	Bayesian Posterior Probability
CAF	Central Analytical Facilities
CAM	Crassulacean Acid Metabolism
CBC	Centre for Biological Control
CBOL	Consortium for the Barcode of Life
COI	Cytochrome c Oxidase Subunit I
DNA	Deoxyribonucleic Acid
EE	Euclidean Error
ERH	Enemy Release Hypothesis
ESS	Effective Sample Size

F1	First Filial Generation
F2	Second Filial Generation
GUI	Graphical User Interface
IA	Identification Accuracy
IAP	Invasive Alien Plant
ISS	Index of Substitution Saturation
ISSc	Critical Index of Substitution Saturation
ISSR	Inter Simple Sequence Repeat
JE	Jaccard Error
KNP	Kruger National Park
matK	Maturase K
MCMC	Markov Chain Monte Carlo
ML	Maximum Likelihood
MRF	Mass Rearing Facility
mtDNA	Mitochondrial DNA
mya	Million Years Ago
NN	Nearest Neighbour
NTA	Non-target Attack
Numt	Nuclear Mitochondrial DNA
ORF	Open Reading Frame
OTU	Operational Taxonomic Unit

PCR	Polymerase Chain Reaction
PGE	Paternal Genome Elimination
RAPD	Random Amplified Polymorphic DNA
rbcL	RuBisCo Large Subunit
RFLP	Restriction Fragment Length Polymorphism
RFU	Relative Fluorescence Units
RNA	Ribonucleic Acid
rRNA	Ribosomal Ribonucleic Acid
SPIDER	SPecies IDentity and Evolution in R
SSR	Simple Sequence Repeat
TBE	Tris/Borate/EDTA
TD-PCR	Touchdown PCR
TID	Threshold Identification
UPGMA	Unweighted Pair Group Method with Arithmetic Mean
<b>Symbols</b>	
$\pi$	Genetic Diversity
ddH <sub>2</sub> O	Double Distilled Water
G <sub>st</sub>	Coefficient of Gene Differentiation
h	Nei's Genetic Diversity
H <sub>s</sub>	Intra-population Diversity
H <sub>t</sub>	Total Genetic Diversity

I Shannon's Information Index

na Number of Alleles

ne Effective Number of Alleles

### **Units**

$\mu$  micro

n nano

# Abstract

The Cactaceae family comprises 15 genera and nearly 2000 species. With one exception, these are all native to the Americas. Numerous cactaceous species are invasive in other parts of the world, resulting in considerable damage to ecosystem functioning and agricultural practices. The most successful biological control agents used to combat invasive Cactaceae belong to the *Dactylopius* genus (Hemiptera: Dactylopiidae), comprising eleven species. The Dactylopiidae are exclusively cactophagous and are usually host-specific. Some intraspecific lineages of dactylopiids, often referred to as 'biotypes', also display host-specificity, and are used to control particular species of invasive Cactaceae. To date, two lineages within *Dactylopius opuntiae* ('ficus' and 'stricta'), and two within *D. tomentosus* ('cholla' and 'imbricata') have been released in South Africa to control *Opuntia ficus-indica* and *O. stricta*, and *Cylindropuntia fulgida* and *C. imbricata*, respectively. The 'californica var. parkeri' lineage is currently under consideration for release in South Africa for the control of *C. pallida*. Australia has already released these five lineages, and approved the release of an additional three in 2017; namely *D. tomentosus* 'bigelovii', 'cylindropuntia sp.', and 'acanthocarpa x echinocarpa'. Many of the *Dactylopius* species are so morphologically similar, and in the case of lineages, identical, that numerous misidentifications have been made in the past. These errors have had serious implications, such as failed attempts at the biological control of cactus weeds. This thesis aimed to generate a multi-locus genetic database to enable the identification of the species and lineages in the Dactylopiidae family, and to test its accuracy. Seven species were included in the analysis, including two lineages within *D. opuntiae* and six within *D. tomentosus*. Genetic characterisation was achieved through the DNA sequencing of three gene regions; namely mitochondrial 12S rRNA and cytochrome c oxidase I (COI), nuclear 18S rRNA, and fragment analysis using two inter-simple sequence repeats (ISSRs). Nucleotide sequences were very effective for species-level identification, where the 12S, 18S, and COI regions showed 100%, 94.59%, and 100% identification accuracy rates, respectively. Additionally, the 12S and COI markers distinguished between half of the *D. tomentosus* lineages ('californica', 'cholla', and 'imbricata'), with identification accuracies of 100%. The 'echinocarpa x acanthocarpa', 'bigelovii', and 'cylindropuntia sp.' lineages formed one clade. None of the DNA genetic markers showed a separation between the 'ficus' and 'stricta' lineages within *D. opuntiae*.

Fragment analysis through the use of ISSRs provided higher-resolution results, and addressed this gap by showing a well-supported separation between the two lineages, and between wild populations collected in the Eastern Cape Province in South Africa. The identification accuracy of the ‘*ficus*’ and ‘*stricta*’ lineages was 81.82%. This is the first time that a method has been developed that can distinguish between these lineages. An additional component of this thesis was the creation of three user-friendly R-based programs to assist with:

1. ISSR data processing.
2. The identification of query *Dactylopius* nucleotide sequences relative to the gene databases created here.
3. A graphical user interface (GUI) version of the R package ‘SPIDER’, which is useful for the assessment of the accuracy of genetic barcode data.

A successful biological control programme relies on the correct identification of the agent in question, and so it is imperative that cactus biological control practitioners are able to distinguish between *Dactylopius* species and lineages in order to release the most effective ones onto target Cactaceae. The laboratory protocols reported, and data processing tools created here, have largely addressed this need and offer valuable practical applications. These include:

1. The flagging of potential new species, cryptic species, and lineages of dactylopiid species released as new biocontrol agents.
2. Validating the identifications made by taxonomists based on morphology.
3. Confirming to which species, and, where applicable, to which lineage, a field-collected sample belongs.
4. Identifying hybrids resulting from lineage crosses.

Ensuring that the correct *Dactylopius* species are utilised for biological control will improve the control of invasive Cactaceae and protect biodiversity and agricultural productivity.

# Contents

<b>1</b>	<b>General introduction</b>	<b>1</b>
1.1	Impacts of Invasive Species . . . . .	3
1.2	Biological control . . . . .	4
1.2.1	Overview . . . . .	4
1.2.2	The biological control pipeline . . . . .	6
	Agent selection and prioritisation . . . . .	6
	Host specificity testing . . . . .	7
	Biological control track record . . . . .	9
1.3	The Cactaceae . . . . .	9
1.3.1	Biological control of invasive Cactaceae . . . . .	11
	<i>Opuntia</i> Mill . . . . .	14
	<i>Cylindropuntia</i> (Engelmann) (Knuth) . . . . .	16
1.4	The Dactylopiidae . . . . .	17
1.4.1	Identification issues . . . . .	18
1.4.2	‘Biotypes’ . . . . .	18
1.4.3	Species summary . . . . .	22
1.5	DNA and molecular barcoding . . . . .	29
1.5.1	History and application . . . . .	29
1.5.2	Caveats . . . . .	30
	Heteroplasmy . . . . .	30
	Maternal inheritance . . . . .	30
	Differential rates of gene evolution . . . . .	31
	Introgression of mitochondrial genes . . . . .	32
	Incomplete lineage sorting . . . . .	32
	Nuclear mitochondrial DNA segments . . . . .	33

1.6	Inter-simple sequence repeats (ISSRs) . . . . .	34
1.6.1	Caveats of inter-simple sequence repeats . . . . .	35
	Homoplasy . . . . .	36
	Error rates . . . . .	36
1.6.2	ISSRs as an identification tool . . . . .	37
1.7	Aims . . . . .	38
<b>2</b>	<b>Materials and Methods</b>	<b>40</b>
2.1	Sourcing of samples . . . . .	40
2.1.1	South Africa . . . . .	40
	Wild populations . . . . .	40
	Uitenhage Mass Rearing Facility . . . . .	40
	<i>Dactylopius tomentosus</i> ‘cholla’ . . . . .	41
	Kruger National Park . . . . .	41
2.1.2	Namibia . . . . .	41
2.1.3	The USA . . . . .	42
2.1.4	Australia . . . . .	42
2.1.5	Saudi Arabia . . . . .	43
2.2	Sample preservation and DNA extraction . . . . .	43
2.3	DNA sequencing . . . . .	43
2.3.1	Gene regions . . . . .	43
2.3.2	Polymerase Chain Reaction (PCR) . . . . .	44
2.3.3	Gel electrophoresis . . . . .	46
2.3.4	Sequence alignment . . . . .	46
2.3.5	Construction of phylogenetic trees . . . . .	47
	Bayesian Inference and Maximum Likelihood . . . . .	47
	Haplotype networks in POPART . . . . .	48
2.3.6	Distance-based testing of barcodes . . . . .	48
	Nearest Neighbour (NN) . . . . .	48
	Threshold ID (TID) . . . . .	48

	Best Close Match (BCM) . . . . .	49
	Scoring of the barcode algorithms . . . . .	49
2.4	Inter-simple Sequence Repeats (ISSRs) . . . . .	50
2.4.1	PCR protocol, data capturing and data processing . . . . .	50
2.4.2	ISSR data processing and analysis . . . . .	50
	Electropherograms to binary data . . . . .	50
	Grouping methods . . . . .	51
	STRUCTURE . . . . .	52
	ISSR error rates . . . . .	53
2.5	Graphical User Interface (GUI) Applications . . . . .	53
2.5.1	R Shiny Applications . . . . .	53
2.5.2	Barcode Testing . . . . .	53
2.5.3	Binary data processing: ‘BINMAT’ . . . . .	54
2.5.4	Identification of query sequences: ‘Dacty-ID’ . . . . .	54
<b>3</b>	<b>Results</b>	<b>55</b>
3.1	Codon saturation . . . . .	55
3.2	Phylogenies . . . . .	55
3.2.1	12S . . . . .	55
3.2.2	18S . . . . .	56
3.2.3	COI . . . . .	56
3.2.4	Concatenation of the 12S and 18S gene regions . . . . .	57
3.3	Sample identification . . . . .	57
3.3.1	South Africa . . . . .	57
3.3.2	Namibia . . . . .	58
3.3.3	The USA . . . . .	58
3.3.4	Australia . . . . .	58
3.4	Identification accuracy . . . . .	65
3.4.1	Nearest Neighbour (NN) results . . . . .	65
3.4.2	12S . . . . .	65

3.4.3	18S . . . . .	65
3.4.4	COI . . . . .	66
3.5	ISSR fragment analysis . . . . .	68
3.5.1	Hierarchical clustering tree . . . . .	68
3.5.2	SplitsTree . . . . .	71
3.5.3	Non-metric Multi-dimensional Scaling (nMDS) Plot . . . . .	72
3.5.4	Structure plots . . . . .	73
3.5.5	Identification accuracy . . . . .	74
<b>4</b>	<b>Discussion</b>	<b>76</b>
4.1	Genetic markers . . . . .	76
4.1.1	12S . . . . .	76
4.1.2	18S . . . . .	77
4.1.3	COI . . . . .	78
4.2	Distance-based barcode testing . . . . .	78
4.3	Posterior probabilities vs bootstrap values . . . . .	79
4.4	ISSR fragment analysis . . . . .	80
4.4.1	‘Stricta’ and ‘ficus’ lineages . . . . .	80
Eastern Cape wild populations . . . . .	80	
Uitenhage Mass Rearing Facility (MRF) . . . . .	81	
The Kruger National Park and Saudi Arabia . . . . .	81	
4.4.2	ISSR identification accuracy . . . . .	81
4.5	Geographical distribution . . . . .	82
4.6	Utility of DNA and molecular barcoding . . . . .	86
4.6.1	New species, cryptic species, and lineages . . . . .	86
4.6.2	Identifying hybrids . . . . .	87
4.6.3	<i>Dactylopius opuntiae</i> : pest or biocontrol agent? . . . . .	87
4.7	Conclusion . . . . .	88
	<b>References</b>	<b>88</b>

<b>5</b>	<b>Appendix A</b>	<b>121</b>
5.1	Extraction protocol . . . . .	121
5.2	DNA sequences . . . . .	121
5.2.1	Submission to Genbank . . . . .	121
5.2.2	Gel images . . . . .	122
5.3	Program settings and file formats . . . . .	129
5.3.1	GeneMarker . . . . .	129
5.3.2	POPART . . . . .	129
5.3.3	SplitsTree . . . . .	130
5.3.4	STRUCTURE . . . . .	131
5.4	Jaccard's transformation . . . . .	132
5.5	Codon saturation . . . . .	132
5.6	ISSR results . . . . .	134
5.6.1	Electropherograms . . . . .	134
5.6.2	Error rates . . . . .	135
5.6.3	Splitstree . . . . .	137
5.6.4	Shepard and scree plots . . . . .	139
5.6.5	ISSR primer statistics . . . . .	140
5.6.6	Genetic distance and diversity . . . . .	140
5.7	R Scripts . . . . .	141
5.7.1	Barcode testing using SPIDER . . . . .	141
5.7.2	Jaccard's transformation . . . . .	142
5.7.3	UPGMA clustering tree . . . . .	143
5.7.4	Non-metric MDS plot . . . . .	143
5.7.5	Three dimensional non-metric MDS plot . . . . .	144
5.7.6	ANOSIM and PERMANOVA . . . . .	145
5.8	Shiny Applications . . . . .	146
5.8.1	Dacty-ID . . . . .	148
5.9	MRBAYES Nexus files . . . . .	150

5.9.1	12S	150
5.9.2	18S	151
5.9.3	COI	152
<b>6</b>	<b>Appendix B: Additional literature review material</b>	<b>153</b>
6.1	The genetic system of the scale insects	153
6.1.1	The evolution of haplodiploidy	155
6.1.2	Paternal genome elimination	157
6.2	Life cycle of the Dactylopiidae	159
6.3	Genetic barcoding application examples	160
6.4	ISSR concept	162

# List of Figures

1.1	Distribution of alien invasive plants in South Africa (Henderson (2001)). . . . .	1
1.2	The invasive range of 57 cactus species, as reviewed by Novoa <i>et al.</i> (2015). . . . .	11
1.3	Gene trees within species trees, illustrating deep coalescence. (1) shows how the most recent common ancestor in the gene tree pre-dates that of the species tree by time T. (2) is another example of how the topology of a gene and species tree can differ. The species tree suggests that A and B share the most recent common ancestor, while the gene tree contrastingly shows that this is true for B and C. (Adapted from Maddison (1997); Nichols (2001); Edwards (2009)).	33
1.4	Potential sources of scoring errors in band-based genotyping methods: A) Determining the optimal threshold value to apply when scoring the presence or absence of a peak. The illustration shows three RFU (Relative Fluorescence Units) threshold levels (a, b and c), the resulting binary output at each, and the consolidated binary output representing a replicate pair (where a 1 and 0 or 0 and 1 outputs a '?', the shared presence of a band outputs a '1' and the shared absence a '0'). B) The effect of bin width on the scoring of peaks. Replicate 1 shows a band of 50.1 base pairs (bp) in size, and replicate 2 shows one of 0.9 bp larger. With a bin width set at b1, both will be scored as present at the same locus, but at a width of b2, these bands will be scored as two separate loci. . . .	38
3.1	Bayesian and Maximum likelihood phylogenies for the 12S gene region. Posterior probability values are above, and bootstrap values are below the branches. Haplotype networks are shown inset for <i>D. opuntiae</i> and <i>D. confusus</i> . Small black dots indicate missing haplotypes, and the diameter of each circle is proportional to the number of genetic sequences that comprise it. Numbers in brackets indicate the number of nucleotide differences between haplotypes.	59
3.2	Bayesian and Maximum likelihood phylogenies for the 18S gene region. Posterior probability values are above, and bootstrap values are below the branches. . . . .	60

3.3	Bayesian and Maximum likelihood phylogenies for the COI gene region. Posterior probabilities are above, and bootstrap values are below the branches. Haplotype networks are shown for <i>D. opuntiae</i> and <i>D. confusus</i> . Small black dots indicate missing haplotypes, and the diameter of each circle is proportional to the number of individuals it comprises. Numbers in brackets indicate the number of nucleotide differences between haplotypes.	61
3.4	Concatenated Bayesian and Maximum Likelihood phylogenies for the 12S and 18S gene regions. Bayesian posterior probability values are above, and Maximum likelihood bootstrap values are below the branches. . . . .	62
3.5	A box plot for the number of fragments amplified by the ISSR 809 and ISSR 826 primers.	68
3.6	Hierarchical clustering tree (UPGMA method) applying Jaccard's transformation for the ISSR binary data obtained from two primers (ISSR 809 and ISSR 826). Approximately unbiased p-values (Au p-value) and the bootstrap probability values (Bp-value) above 70 are shown as black dots above and below the relevant clades. One thousand bootstrap repetitions were run. MRF = Uitenhage Mass Rearing Facility. . . . .	70
3.7	SplitsTree graphical output (NeighborNet method, applying Jaccard's index to calculate the distance matrix) for <i>Dactylopius opuntiae</i> individuals (ISSR 809 and ISSR 826). The Uitenhage and Namibia clusters are known to be the 'ficus' lineage, and the Kruger National Park (KNP) contain the 'stricta' lineage. Individuals in the 'wild populations' group were sampled from <i>Opuntia ficus-indica</i> and <i>O. engelmannii</i> host plants in the Eastern Cape. MRF = the Mass Rearing Facility at Uitenhage (Eastern Cape). Bootstrap values above 75 are shown. Bold names in the KNP group denote the known 'stricta' source population samples (kept by Hildegard Klein and John Hoffmann). . . . .	71
3.8	A non-metric MDS plot based on the binary data (applying Jaccard's distance transformation) obtained from ISSR 809 and ISSR 826 to represent two different <i>Dactylopius opuntiae</i> lineages. Blue circles = 'ficus', red squares = 'stricta', and green triangles = the Uitenhage Mass Rearing Facility. . . . .	72
3.9	STRUCTURE results using Bayesian analysis (MCMC = 50 000, burn-in = 100 000), showing genetic clustering patterns for <i>Dactylopius opuntiae</i> individuals. Results for both $K = 2$ and $K = 5$ are shown for runs with and without priors. . . . .	75

4.1	A comparison of the phylogenies produced by A) Rodríguez <i>et al.</i> (2001) based on morphological characteristics, Ramírez-Puebla <i>et al.</i> (2010) based on the B) 12S and C) 18S rRNA gene regions, and D) the present study, represented as a cladogram for the 18S region (not to scale). The values on the branches of diagrams B, C, and D are Bayesian posterior probability values. NA = North America; SA = South America. . . . .	85
A1	12S and 18S PCR products visualised on 1% agarose gel. Note the double bands in many of the 12S products; at the 500 and 1500 base pair mark. . . . .	122
A2	ISSR banding pattern on a 1.5% agarose gel for the two <i>D. tomentosus</i> lineages ('cholla' and 'imbricata'), run as a preliminary test before fragment analysis. Ladders are on the far left and right, marked at 100, 500 and 1500 base pair sizes. . . . .	122
A3	12S rRNA Bayesian phylogeny based on the alignment obtained from R-Coffee, which aligns RNA sequences according to their secondary structure. The HKY + I + G model was used, where shapepr = 0.9980, pinvarp = 0.1420 and tratiopr = 1.4096. State frequencies were 0.5155, 0.0897, 0.0282 and 0.3665. Major clades have been collapsed. . . . .	128
A4	An example of Jaccard's similarity index calculation. Note that the shared absence of bands are not taken into account in the calculation of pairwise distances. Jaccard's distance = $1 - J_1$ , and therefore 0.43 in this case. . . . .	132
A5	Saturation graphs for codon positions 1 to 3 for the COI gene region; showing the relationship between genetic distance and the rate of transitions (s) and transversions (v). 133	
A6	An overlay view of fragment sizes between 300 and 380 bp for samples (1) <i>Dactylopius opuntiae</i> 'stricta', (2) <i>D. austrinus</i> , (3) <i>D. opuntiae</i> 'ficus-indica', (4) <i>D. tomentosus</i> 'cholla', (5) <i>D. confusus</i> , (6) <i>D. tomentosus</i> 'pallida' and (7) <i>D. opuntiae</i> collected from Flagstaff (AZ, USA). The y-axis represents Relative Fluorescence Units (RFU), and the x-axis displays fragment size in base pairs. . . . .	134
A7	Euclidean and Jaccard error rates for ISSR primers 826 and 809, showing the proportion of samples greater than a particular error rate. Values are representative of the full data set. 135	
A8	Graphical representations of the effect of the removal of samples (ISSR 809 and ISSR 826) with a Jaccard error rate greater than a specific threshold on A) the average Jaccard and Euclidean error rates and B) the proportion of removed samples from the data set. . .	136

A9	Splitstree diagram representative of ISSR 809 only. Mass Rearing Facility (MRF) samples are shown in blue. KNP = Kruger National Park, and source pops. = the ‘stricta’ culture kept by Hildegard Klein and John Hoffmann. . . . .	137
A10	Splitstree diagram representative of ISSR 826 only. Mass Rearing Facility (MRF) samples are shown in blue. KNP = Kruger National Park, and source pops. = the ‘stricta’ culture kept by Hildegard Klein and John Hoffmann. . . . .	138
A11	A) Scree plot showing the relationship between the number of dimensions used for a non-metric MDS, and the corresponding stress score. Stress scores below 15% are considered to be a good fit for the data (indicated by the dashed red line). Stress scores for $K = 2$ and $K = 3$ are indicated by the dashed green lines. Shepard plots are shown in B) ( $K = 2$ ) and C) ( $K = 3$ ), plotting the relationship between the original distance data and the distances resulting from the ordination transformation. A high $R^2$ value is desirable, as that indicates an accurate reflection of the data in the ordination plot. . . . .	139
A12	Box plots for within and overall group genetic distances for <i>D. opuntiae</i> , measured by the Jaccard distance (1-J). A) ‘ficus’ taken as a whole and B) wild populations removed from ‘ficus’ and put into its own group. Solid black dots are the means, and the same letters above each box indicate non-significance. . . . .	141
A13	Genetic barcode tester GUI showing the components of the application. Numbers in circles indicate the order in which data is uploaded and processed. . . . .	146
A14	Binary matrix processing application (‘BINMAT’), where the user uploads a .csv file containing binary data replicates in order to consolidate them. Summary statistics, error rates and the construction of a UPGMA clustering tree are also available. Non-metric MDS plots can be created by uploading a consolidated binary matrix, where filtering of samples with a low peak count can be applied. . . . .	147
A15	Identification GUI application, where the user can test the identity of a query sequence against a preselected genetic database. A phylogenetic tree is displayed on the screen, with the position of the query sequence highlighted in red. . . . .	149
B1	Difference between normal and inverted meiosis. Adapted from Brown and Nur (1964); Wrench <i>et al.</i> (1994); Bongiorno <i>et al.</i> (2004) . . . . .	154

B2	Paternal genome elimination in scale insects, showing the lecanoid, diaspidid and Comstock- iella genetic systems. Adapted from Bull (1979); Burt and Trivers (2009); Ross <i>et al.</i> (2010b)	158
B3	Lifecycle of <i>Dactylopius</i> . Adapted from Moran and Cabby (1979) . . . . .	159
B4	An illustration of the concept of ISSR-PCR showing the use of 3' and 5'-anchored primers. A 5-anchored primer attaches to the flanking region of a microsatellite, where a 3-anchored primer attaches to the microsatellite region flanking the ISSR and forms a shorter PCR product. The ISSR region varies in size at different loci in the genome, producing unique banding patterns for each unique individual. Diagram adapted from Zietkiewicz <i>et al.</i> (1994); Reddy <i>et al.</i> (2002); Ng and Tan (2015). . . . .	162

# List of Tables

1.1	National Environmental Management: Biodiversity Act (NEM:BA) Category 1 and 2 invasive cactus species in South Africa. Category 1a and 1b weeds require compulsory control, and need to be removed and destroyed. Category 2 weeds are regulated by area, and require demarcation permits (none are, however, allowed in riparian zones) (Department of Environmental Affairs, 2016). . . . .	13
1.2	List of the Cactaceae in the Australian Weeds of National Significance (WoNS) database (Sheehan and Potter, 2017). . . . .	14
1.3	A list of the 11 recorded <i>Dactylopius</i> species and intraspecific lineages . . . . .	17
2.1	Primers used for the amplification of the COI, 12S, and 18S gene regions in <i>Dactylopius</i> species.	44
2.2	PCR protocol for the 12S, 18S, COI-A, and COI-B genes in <i>Dactylopius</i> . . . . .	45
2.3	PCR protocol for the 12S, 18S, COI-A, and COI-B genes. . . . .	45
2.4	12S Touchdown-PCR protocol that addressed double-band amplification. . . . .	46
2.5	PCR protocol for the ISSR 809 and ISSR 826 primers for <i>Dactylopius</i> spp. . . . .	50
3.1	Within and between group p-distance matrix for the 12S gene, using the Kimura 2-parameter model. Species are represented by a total of 151 nucleotide sequences consisting of 779 positions. Bold values along the diagonal represent within-group p-distances. . .	63
3.2	Within and between group p-distance matrix for the 18S gene, using the Kimura 2-parameter model. Species are represented by a total of 153 nucleotide sequences consisting of 630 positions. Bold values along the diagonal represent within-group p-distances. . .	63
3.3	Within and between group p-distance matrix for the COI-A and COI-B regions, using the Kimura 2-parameter model. Species are represented by a total of 95 nucleotide sequences consisting of 603 positions. Bold values along the diagonal represent within-group p-distances.	64

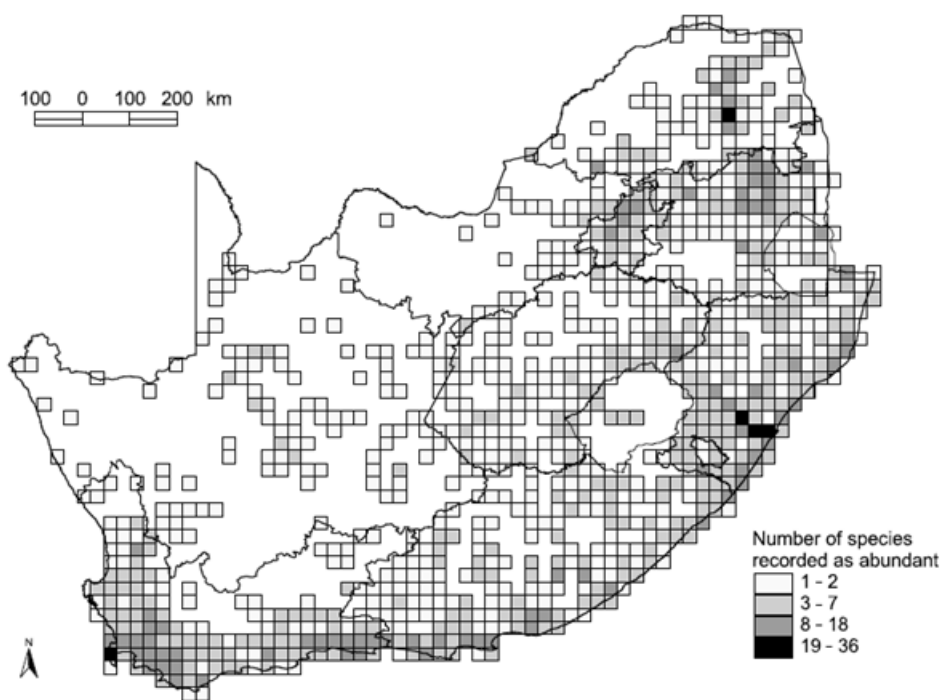
3.4	Results of the Best Close Match (BCM), Threshold ID (TID) and Nearest Neighbour (NN) barcode testing algorithms for the 12S, 18S and COI gene regions. Values for BCM and TID are shown at the default 1% and at the optimum threshold value. Results are shown at the species and lineage level. In the COI section, <i>D. tomentosus</i> (G) indicates a test conducted where the ‘bigelovii’, ‘cylindropuntia sp.’, and ‘echinocarpa x acanthocarpa’ sequences were grouped as one lineage. . . . .	67
3.5	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. . . . .	68
3.6	ANOSIM intra-group statistics for <i>D. opuntiae</i> ‘stricta’, ‘ficus’ and Mass Rearing Facility (MRF) groups, showing their R values, followed by uncorrected p-values in brackets. Permutations = 999. Overall R value = 0.44, p = 0.001. Asterisks denote significance (p < 0.05). . . . .	73
3.7	ANOSIM intra-group statistics for <i>D. opuntiae</i> , where the ‘ficus’ group is split into wild populations and the ‘ficus’ collected from around Uitenhage and in Namibia. R values are shown, followed by uncorrected p-values in brackets. Permutations = 999. Overall R value = 0.72, p = 0.001. Asterisks denote significance (p < 0.05). . . . .	73
3.8	Permutational Multivariate Analysis Of Variance Using Distance Matrices for <i>D. opuntiae</i> ‘stricta’, ‘ficus’ and Mass Rearing Facility (MRF) groups. Pairwise p-values are shown. Permutations = 999, DF = 2, F = 6.60, R <sup>2</sup> = 0.22, p = 0.001. Asterisks denote significance (p < 0.05). . . . .	73
3.9	Permutational Multivariate Analysis Of Variance Using Distance Matrices intra-group statistics for <i>D. opuntiae</i> where the ‘ficus’ group is split into wild populations and the ‘ficus’ collected from around Uitenhage and in Namibia. MRF = Mass Rearing Facility. Pairwise p-values are shown. Permutations = 999, DF = 3, F = 11.69, R <sup>2</sup> = 0.43, p = 0.001. Asterisks denote significance (p < 0.05). . . . .	73
3.10	Barcode testing results for the Best Close Match (BCM), Threshold ID (TID), and Nearest Neighbour (NN) tests, at the species and lineage level. . . . .	74

A1	Genbank accession numbers for 12S, 18S and COI genetic sequences; with the extraction sample ID, species, lineage, locality and host plant information. CT Bot. Gard. = Cape Town Botanical Gardens, EC = Eastern Cape, KNP = Kruger National Park, MRF = Mass Rearing Facility. Black dots indicate samples collected in the native range in 2017.	123
A2	Sample IDs for additional DNA extractions used for ISSR analyses. . . . .	126
A3	Genbank sequences used to supplement the dataset for this project. PcoF1 & LepR1 and DTOMf & HCO2198 are both COI primers that amplify the same region of the COI gene.	127
A4	Rhodes University and Agricultural Research Council (ARC) accession numbers for <i>Dactylopius</i> samples from the native range (USA), identified by Ian Millar in 2019. Representative sample IDs refer to those used in the gene phylogenies in this project. See Table A1 for the full list of sequences. . . . .	127
A5	Codon saturation results from DAMBE for the protein-coding COI-A and COI-B gene regions. A significant p-value and an Iss value < the Iss.cSym value means that there is little saturation present. . . . .	132
A6	POPGENE statistical output for each individual primer and their concatenation for the ‘stricta’ and ‘ficus’ genetic clusters. na = number of alleles, ne = effective number of alleles, h = Nei’s genetic diversity, I = Shannon’s information index, No. PM loci = number of polymorphic loci, % PM loci = percentage polymorphic loci, Ht = total genetic diversity, Hs = intra population genetic diversity, Gst = coefficient of gene differentiaion. All values presented are averages. For the intra-group statistics, Saudi Arabia + Kruger N. Park + Klein & Hoffmann = ‘stricta’, and Uitenhage + Namibia + wild populations = ‘ficus’.	140
A7	POPGENE statistical output for the concatenation of ISSR 809 and ISSR 826 for the ‘stricta’ and ‘ficus’ genetic clusters. na = number of alleles, ne = effective number of alleles, h = Nei’s genetic diversity, I = Shannon’s information index, No. PM loci = number of polymorphic loci, % PM loci = percentage polymorphic loci, Ht = total genetic diversity, Hs = intra population genetic diversity, Gst = coefficient of gene differentiaion. All values presented are averages. For the intra-group statistics, Saudi Arabia + Kruger N. Park + Klein & Hoffmann = ‘stricta’, and Uitenhage + Namibia + wild populations = ‘ficus’.	141
B1	Cross-disciplinary examples of genetic barcoding . . . . .	160

# Chapter 1

## General introduction

Invasive alien plant (IAP) invasions in South Africa date back to the 1600s with the advent of European trade and settlement in the Cape (van Wilgen *et al.*, 2001; Zimmermann *et al.*, 2004; Moran *et al.*, 2005; Alston and Richardson, 2006). It is now home to over 180 IAP species and is one of the most heavily invaded countries in the world (Henderson, 2001; Richardson and van Wilgen, 2004) (Figure 1.1). Australia is in a very similar situation. Of the approximately 3207 alien plant species that have established in the country, about 500 taxa have become invasive; costing the economy approximately AUD\$4 billion annually (Invasive Plants and Animals Committee, 2017).



**Figure 1.1:** Distribution of alien invasive plants in South Africa (Henderson (2001)).

A 1996/7 survey conducted by Le Maitre *et al.* (2000) estimated that IAPs occupied approximately 10.1 million ha (or 1.736 million condensed ha) of South Africa. Combined estimates published 14 years later by Kotzé *et al.* (2010) and van den Berg (2010) raised that figure to 1.813 million condensed ha. The annual economic cost associated with the damage inflicted by these IAPs was estimated to be approximately R6.5 billion (at 2008 inflation values) (de Lange and van Wilgen, 2010). Water loss and the degradation of grazing land were the highest contributors to this figure.

One of the major groups of IAPs in both South Africa and Australia are the Cactaceae (Moran and Zimmermann, 1991a; Zimmermann *et al.*, 2009; Klein, 2011; Paterson *et al.*, 2011). The first successful biological control initiative on record for a weed was for the control of *Opuntia monacantha* Haworth in India in the 1800s, and was the result of a misidentification of a *Dactylopius* species (Greenfield, 2005; Zimmermann *et al.*, 2009). *Dactylopius ceylonicus* Green (Hemiptera: Dactylopiidae) was initially introduced to India in the belief that it was *D. coccus*; a species reared for its high output of carminic acid used for the production of red dye. The insect caused such damage to *O. monacantha* that the dye industry failed, but the insect continued to be used as a control agent for *O. monacantha* infestations in the country, and in Sri Lanka (Zimmermann *et al.*, 2009). Following its success in India, South Africa released *D. ceylonicus* to combat *O. monacantha* in 1913, which resulted in complete control of the weed within a few years and is the first intentional biological control programme on record globally (Lounsbury, 1915; Zimmermann *et al.*, 2004; Paterson *et al.*, 2011). Australia followed suit and released *D. ceylonicus* in 1914, which established successfully and delivered widespread control of the same weed (Winston *et al.*, 2014). The Dactylopiidae family (*Dactylopius* Costa) is predominantly host specific to cactaceous species in the *Opuntia* and *Nopalea* genera (De Lotto, 1974; Campana *et al.*, 2015; Van Dam *et al.*, 2015). These two cactus genera are genetically and taxonomically closely related (Griffith and Porter, 2009). The Dactylopiidae family is monogeneric and comprises eleven species of cactophagous insects (commonly referred to as ‘cochineal’) (Campana *et al.*, 2015; Van Dam *et al.*, 2015). Four *Dactylopius* species are used in South Africa and Australia as biocontrol agents for invasive *Opuntia* and *Cylindropuntia* (Engelm.(Kuth)) cacti (Zachariades, 2018). Twelve other countries around the world have released *Dactylopius* species to control invasive cacti, where the impacts recorded are generally very high when establishment was successful (Winston *et al.*, 2014). Many of the *Dactylopius* species contain distinct genetic groups that display differential host specificities (Volchansky *et al.*, 1999; Mathenge *et al.*, 2009a, 2010a,b; 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), this thesis will instead use the term ‘lineage’. This is discussed in detail in Section 1.4.2. 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 methods. Being able to identify the species and lineages of *Dactylopius* insects is fundamental to the practice of biological control of Cactaceae worldwide.

## 1.1 Impacts of Invasive Species

Alien invasive organisms have the potential to significantly alter novel environments and affect native species at the genetic through to the ecosystem level (Thompson, 1991; Crooks, 2002; Lockwood *et al.*, 2013). Vilà and Hulme (2017) grouped the impacts of invasive species on ecosystem services into four categories; namely that of 1) supporting, 2) provisioning, 3) regulating, and 4) cultural and human well-being.

Supporting services include important ecosystem processes incorporating nutrient and energy flows, such as carbon sequestration and primary production, nutrient cycling, habitat structure, and hydrology. As an example, invasive *Acacia mearnsii* De Wild. (black wattle) dominate riparian zones in South Africa and use significantly more water than indigenous plants. This causes major reductions in water catchment yields and streamflow (Le Maitre *et al.*, 2000; Dye and Jarman, 2004). Le Maitre *et al.* (2000) estimated that IAPs in South Africa use approximately 3.3 billion m<sup>3</sup> of water annually. This is a very disconcerting figure taking into account that many parts of the country face severe water shortages.

Provisioning services include the resources humans need for food, fuel, water, and other essential materials such as fibre (Pejchar and Mooney, 2009; Vilà and Hulme, 2017). Approximately 10% of the Earth’s land surface comprises arable land used for agriculture and horticulture (Fried *et al.*, 2017). These areas are particularly vulnerable to biological invasions due to the high incidence of fertilisation and disturbance, where weeds can interfere through competitive, allelopathic, and parasitic mechanisms (Fried *et al.*, 2017). *Parthenium hysterophorus* L. (famine weed), for example, has invaded agricultural lands in parts of Africa, Asia, and Australia (McConnachie *et al.*, 2011). The weed is highly competitive and allelopathic, and releases phenolics and sesquiterpenes into the surrounding soil that inhibits the growth of important crops such as cereals, vegetables, and grasses (Evans, 1997).

Regulating services are those that sustain and adjust an ecosystem, such as pollination services, disturbance regimes, food webs, water purification, soil stabilization, climate regulation, and the role of endemic species (e.g. mutualistic relationships) (Pejchar and Mooney, 2009; Vilà and Hulme, 2017). As an example of how a disturbance regime can be affected, invasive plants can alter fire regimes by serving as a fuel source and promoting more intense fires (D’Antonio and Vitousek, 1992; Brooks *et al.*, 2004; Milton, 2004; Le Maitre *et al.*, 2014). This allows invaders to outcompete native species that are not suitably adapted to this disturbance and do not recover as quickly (Milton, 2004).

Cultural and human well-being services entail the aesthetics of nature and the spread of diseases (Vilà and Hulme, 2017). For example, *Eichhornia crassipes* infestations in fresh water bodies limits boating activity and other water recreation activities (Villamagna and Murphy, 2010). Invaded areas tend to display lower diversity, where invasive species tend to dominate the landscape and reduce its aesthetic value (Kueffer and Kull, 2017). Invasive species can act as vectors and reservoirs of diseases, such as *Aedes aegypti* L. (‘yellow fever mosquito’) and *A. albopictus* Skuse (‘Asian tiger mosquito’) (Diptera: Culicidae) that vector the Chikungunya virus, and *Phlebotomus* Loew and *Lutzomyia* Frana spp. (‘sandflies’) (Diptera: Psychodidae) that vector Leishmania pathogens (Rabitsch *et al.*, 2017).

The impacts associated with invasive Cactaceae touch on all four invasion impact categories, predominantly relating to the loss of biodiversity, ecological functioning, and agricultural output (Kaplan *et al.*, 2017; Novoa *et al.*, 2019). Areas of dense cactus infestation leads to the exclusion of indigenous plants and animals through competitive exclusion and damage caused by spines and glochids (Walters *et al.*, 2011). Negative impacts on agricultural practices are caused by reduced grazing areas for livestock and the resulting decrease in productivity, costs incurred by mechanical and chemical control practices (where the costs of control sometimes exceed the value of the land itself), injuries to livestock caused by spines, and damage to fleece and hides (Lloyd and Reeves, 2014).

## 1.2 Biological control

### 1.2.1 Overview

van Driesche *et al.* (2009) defines biological control as “*The use of populations of natural enemies to suppress pest populations to lower densities, either permanently or temporarily*”. Biological control is a

broad term encompassing the following methods:

1. ‘Classical’, which is the most common approach, and refers to the use of natural enemies from the native range of a target pest (which share an evolutionary history) to achieve control.
2. ‘Neo-classical’ or ‘new association’ methods, which makes use of non-native agents to control native species (which share no previous evolutionary history).
3. ‘Augmentative’, where the natural enemies of a target pest are mass-reared and released in order to augment naturally occurring enemy populations.
4. ‘Conservation’, which focuses on the maintenance of natural enemy populations in order to optimise their efficacy as parasitoids, pathogens, or predators.

(McFadyen, 1998; Collier and Van Steenwyk, 2004; van Driesche *et al.*, 2009; Gullan and Cranston, 2014; Begg *et al.*, 2017).

Classical biological control is based primarily upon the enemy release hypothesis (ERH), which proposes that when a plant species is introduced into a new environment, it is released from its natural enemies (particularly specialist enemies) and is more readily able to spread quickly and invest resources into growth and reproduction (Keane and Crawley, 2002; McEvoy, 2002; Liu and Stiling, 2006). This often involves energy trade-offs where chemical and structural defenses are down-regulated in favour of rapid growth and seed production (Blossey and Notzold, 1995; Van Dam *et al.*, 1996). The ERH also predicts that it will be rare for the natural enemies of native plants to switch hosts and begin feeding on the invader, and that native generalist enemies will be more likely to feed on native species (Keane and Crawley, 2002).

The benefits of biological control include:

1. Permanency of control and the self-perpetuation of agent populations.
2. Host-specificity with little chance of non-target effects.
3. High cost-to-benefit ratios relative to mechanical and chemical control methods.
4. Being environmentally friendly (involving no toxic residues, and requiring a low energy input).
5. Being able to naturally spread across the range of the host plant.
6. The ability to be used in combination with other control methods as part of an integrated pest management (IPM) programme.

(Greathead, 1995; van Wilgen and Lange, 2004; Culliney, 2005; van Driesche *et al.*, 2009; Zachariades

*et al.*, 2017).

## 1.2.2 The biological control pipeline

A review by McFadyen (1998) summarised the steps involved in a biological control programme as follows: exploration in the weed's region/s of origin → selection of potential agents → host specificity testing → mass rearing of the chosen agent/s → releases into target areas → post-release evaluation and the implementation of integrated control strategies. The two central steps are that of agent selection and host specificity testing, which will briefly be reviewed here.

### Agent selection and prioritisation

There is often an array of phytophagous species associated with a particular host plant (Frost, 1954; Southwood, 1961), but they do not all inflict the same mode and/or level of damage. Wilson (1961), for example, reported that only five of the 52 insects imported for the control of *Opuntia ficus-indica* in Australia could serve as damaging agents. The costs and time required for testing can quickly become unfeasible and impractical (Harris, 1991; McFadyen, 1998; Paterson *et al.*, 2014). Host specificity testing is estimated to take approximately three scientist years and AUD\$460 000 (at 1997 prices) per agent (McFadyen, 1998). The prioritisation of agents is thus integral to a biological control programme, such that only the most effective candidates are selected to deliver the highest level of targeted damage possible. This concept is central to this thesis, as the identification of the correct *Dactylopius* lineage is necessary to inflict the most damage to the target cactus species. Quoting McFadyen (1998), predicting the success of an agent prior to release and selecting the most effective one is the 'holy grail of weed biocontrol'. A number of factors need to be considered in order to select the most suitable candidate, including the following from a general scoring system originally proposed by Harris (1973) and modified by Goeden (1983):

1. The degree of host-specificity displayed (the agent must express suitable levels of specialisation for release in the introduced range).
2. The level of direct and indirect damage caused to the host plant (where damage to vascular and mechanical support tissue is the most effective).
3. Phenology of attack (where a prolonged attack is optimal that extends over the growing and/or reproductive phase of the target weed)

4. The number of generations and progeny per generation produced (where a multivoltine and highly fecund agent is desired).
5. Extrinsic mortality factors (where ideally only specialised natural enemies are associated with the agent, and competition by other species for the host plant is negligible).
6. Feeding behaviour (where gregarious feeding is desired for greater damage).
7. Compatibility with other agents.
8. Distribution (where ideally the agent covers a large part of the range of the target weed).
9. Previous successes as a biological control agent in other parts of the world.

Although the proposed scoring system was criticised for not being thorough enough (Wapshere, 1985; Cullen, 1995; Kriticos, 2003), it presented an initial framework and opened further discussion. Cullen (1995), Kriticos (2003), and Van Klinken and Raghu (2006) suggest that a more integrative approach to agent selection should be based on ecological modelling. This involves understanding the interactions between the agent and its host plant, and being able to predict how various factors such as disturbance regimes, competition, plant succession, different life stages of the host plant, and climatic variation may influence their dynamics (McEvoy and Coombs, 1999). Closely linked to the concept of ecological modelling is that of climatic and genotypic matching; where the climatic compatibility between the native and introduced range is assessed, and the genetic variation (i.e. the number of haplotypes present) in both the host plant and agent is considered (Iline and Phillips, 2004; Dhileepan *et al.*, 2006; Goolsby *et al.*, 2006; Robertson *et al.*, 2008; Paterson *et al.*, 2009; van Driesche *et al.*, 2009; Paterson *et al.*, 2014; Fischbein *et al.*, 2019; Jourdan *et al.*, 2019; Sutton, 2019). Despite all the tools available, extrapolating results from a controlled environment to field conditions remains a major challenge (see for example Buccellato *et al.* (2019)).

### **Host specificity testing**

The safety of a biological control programme is of vital importance, and requires a thorough risk assessment of the potential non-target effects that could arise following the release of an agent into a target area (McEvoy, 1996; Secord and Kareiva, 1996; McFadyen, 1998; Heard, 1999; Van Klinken, 1999; Schaffner, 2001; Heard, 2002; van Driesche *et al.*, 2009). This form of risk assessment is referred to as ‘host specificity testing’, and aims to elucidate what the agent’s host range breadth is, and its relative acceptability of different host plant species (Van Klinken, 1999). Van Klinken (1999) outline three central steps in the testing process, namely:

1. Identifying which stages in an agent's life history need to be host-specific (larval instars might display differential feeding behaviours, and adults might damage non-target plants through feeding, oviposition, or the transferal of pathogens).
2. Estimating the fundamental host range, which is the range of host plants on which the agent is capable of completing its life cycle (Schaffner, 2001). This is tested through the use of the centrifugal-phylogenetic method proposed by Wapshere (1974). The method entails the testing of an array of plant species in the agent's host range; beginning with those most closely related to the target weed, working outwards to more distantly related species. Plants occurring in the same habitat, that share similar chemical and/or morphological traits, or serve as hosts to congeneric insects to that of the target weed are included in the testing procedure. Importantly, economically important crops and rare and endangered species in the same family as the target weed are prioritised. The testing procedure encompasses choice and no-choice tests (Schaffner, 2001), continuation, oogenesis, and open-field tests (van Driesche *et al.*, 2009; Schaffner *et al.*, 2018).
3. Extrapolating to the field, which involves predicting what the realised host range will be under field conditions. The realised host range comprises the host plants that the agent can develop on under field conditions, and is initially predicted by taking surveys of related host plants occurring sympatrically with the target weed (Schaffner, 2001), followed by open field tests if possible (Balciunas *et al.*, 1996).

In cases where test results indicate that an agent will cause damage to one or more non-target species, a cost to benefit assessment is required to determine whether the relative value of the non-target species is higher than the damage caused by the target weed (McFadyen, 1998; van Driesche *et al.*, 2009). Cost to benefit analyses have been used in the past to determine whether an agent should be released (McFadyen and Marohasy, 1990; Olckers *et al.*, 1995; Cristofaro *et al.*, 1998; Haye *et al.*, 2006)

Risk assessments are important to determine potential non-target effects associated with a released agent (Paynter *et al.*, 2015, 2018; Hinz *et al.*, 2019). This is done by quantifying the performance of an agent on target and non-target plants, and calculating risk scores that can predict the probability of attack (Paynter *et al.*, 2015).

## Biological control track record

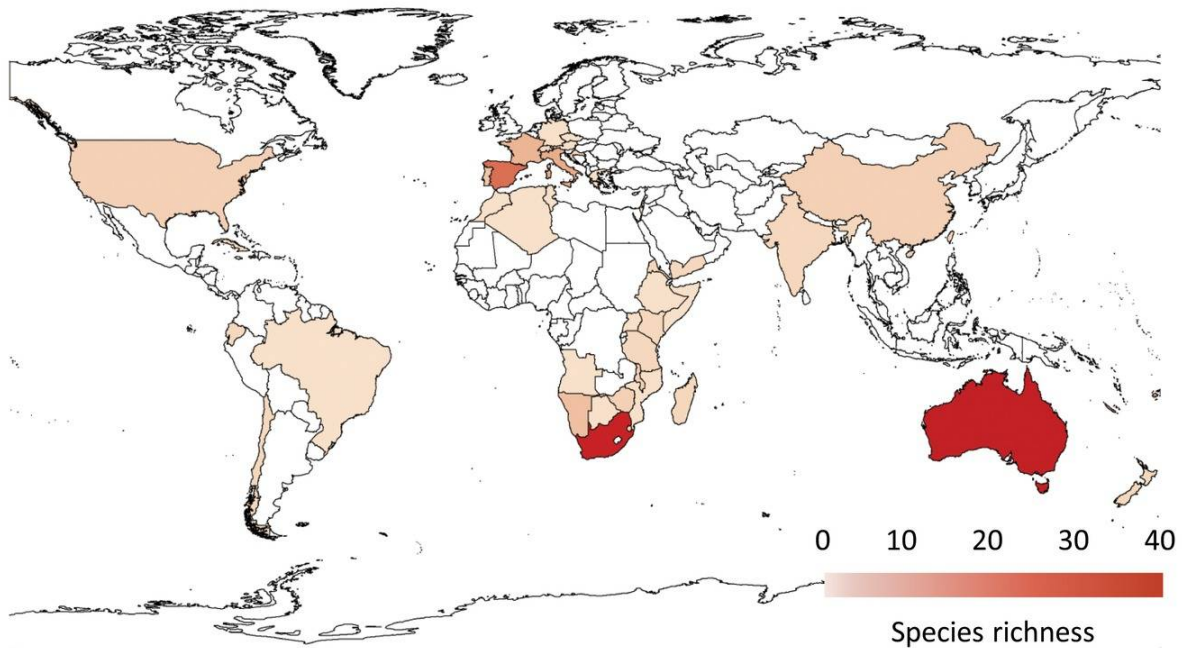
Using the data compiled by Winston *et al.* (2014), Schwarzländer *et al.* (2018) summarised that up to the year 2012, 468 biological control agent species have been released onto 175 weed species (comprising 48 plant families) across the world. Of the agents released, 70.9% established successfully. Post-release surveys determined that only 7% of releases resulted in no impact at all on the target weed, and that 55% of agents caused heavy, medium, or variable impact in at least one release. Of the targeted 175 weed species, 115 (65.7%) have been controlled to some degree by at least one biological control agent.

A review by Hinz *et al.* (2019) stated that of the 457 species of biological control agents intentionally released between the years 1863 and 2008, 60 (13.1%) were associated with non-target attacks (NTAs). Of those 60 NTAs, 42 involved attacks on native species. In agreement with Sheppard *et al.* (2003) and Suckling and Sforza (2014), the review concluded that less than 1% of all weed biological control releases have the potential to result in negative non-target effects. Additionally, it found that the proportion of agent releases associated with NTAs has decreased despite the increase in releases over the years. This is most likely due to a combination of more stringent host specificity testing measures and stricter regulations regarding the release of biocontrol agents. Hinz *et al.* (2019) emphasise that post-release monitoring should be as important as pre-release host specificity testing in order to accurately quantify agent impact and NTAs.

## 1.3 The Cactaceae

There are approximately 130 genera and 2000 species in the Cactaceae family (order Caryophyllales); all of which are indigenous to the Americas (Anderson, 2001; Wallace and Gibson, 2002; Zimmermann *et al.*, 2009; Novoa *et al.*, 2015; Christenhusz and Byng, 2016). The one exception is the epiphyte *Rhipsalis baccifera* Stern ('mistletoe cactus'), which is possibly native to southern Africa and Madagascar (Wallace and Gibson, 2002; Zimmermann *et al.*, 2009). The centre of cactus diversity in the native range is in Mexico, where the three most speciose genera are *Opuntia*, *Echinopsis*, and *Mammillaria* (Novoa *et al.*, 2015). The Cactaceae family evolved about 65 mya in South America (Anderson, 2001), and radiated about 30 to 35 mya in response to the aridification of the Americas and increased atmospheric CO<sub>2</sub> (Arakaki *et al.*, 2011; Majure *et al.*, 2012). The Cactaceae comprise four subfamilies; namely the Pereskioideae, Cactoideae, Maihuenioideae, and Opuntioideae (Labra *et al.*, 2003). Their distribution spans from southern Patagonia in Argentina to British

Columbia in Canada (Anderson, 2001; Edwards *et al.*, 2005). Most cactaceous plants are succulents and are adapted to extreme xeric environments, although some epiphytes have even been found to thrive in rain forests (Anderson, 2001). Most species possess leafless photosynthetic stems and protective spines, which are modified leaves arising from areoles (Eggli, 1993). The spines of cactaceous plants are adapted for a number of functions across different species, such as protection from herbivores, camouflage, thermoregulation, water conservation, and dispersal (Anderson, 2001). The Cactaceae display a diverse range of growth forms and sizes, from the smallest species (*Blossfeldia liliputana* Werdermann) measuring 9 mm in diameter to one of the largest that stands nearly 20 m tall (*Pachycereus pringlei* Britton & Rose, the ‘Giant Mexican cardon’) (Anderson, 2001). Some botanists have classified 13 different growth forms within this family, ranging from globular and columnar, to leaf-like and clustering (Anderson, 2001). Cactaceous plants typically display a diploid chromosome system of  $2n = 22$ , but numerous species in the Opuntioideae family are known to display polyploidy (Anderson, 2001). Additionally, asexual reproduction is common in many Cactaceae, where stem joints can break off the main plant, grow adventitious roots, and give rise to new plants (Anderson, 2001). Although the Cactaceae are endemic to the New World, they have become naturalised widely outside their native distribution, and, in many cases, have become serious invaders (Novoa *et al.*, 2015). Of the 1922 cactus species reviewed by Novoa *et al.* (2015), 57 (3%) were invasive in other parts of the world (Fig. 1.2). The review reported that *O. ficus-indica* was the most widespread invader; having been recorded in 22 countries outside its native range. The countries most heavily invaded by cactaceous weeds are Australia, South Africa, and Spain; with 39, 35, and 24 recorded species, respectively.



**Figure 1.2:** The invasive range of 57 cactus species, as reviewed by Novoa *et al.* (2015).

### 1.3.1 Biological control of invasive Cactaceae

The two predominant countries implementing cactus biocontrol are South Africa and Australia (Moran and Zimmermann, 1991b; Paterson *et al.*, 2019). Numerous cactus species were introduced to these countries as ornamental and hedge plants, as a source of fruit and fodder, and for the establishment of a red dye industry during their early colonial periods (Zimmermann *et al.*, 2009; Paterson *et al.*, 2011; Kaplan *et al.*, 2017; Sheehan and Potter, 2017) (Tables 1.1 and 1.2). There are four main groups of invasive cacti: namely the ‘platyopuntia’ (e.g. *Opuntia ficus-indica*, *O. stricta*, and *O. monacantha*), ‘cylindropuntia’ (e.g. *Cylindropuntia imbricata*), ‘organ-pipe’ (e.g. *Harrisia martinii* and *Cereus jamacaru*), and ‘leafy’ (e.g. *Pereskia aculeata*) (Klein, 2002). The main adaptations enabling the invasibility of these plants are their:

1. Tolerance to xeric conditions and their water-efficient CAM (Crassulacean Acid Metabolism) photosynthetic strategy.
2. Ability to outcompete native vegetation under disturbed conditions.
3. Ability to reproduce both sexually and asexually.

(Zimmermann *et al.*, 2009).

Approximately 400 cactus taxa have been introduced to South Africa (Kaplan *et al.*, 2017), of which

35 species are invasive (Novoa *et al.*, 2015). Twenty three of these have been subject to biological control using 19 insect species (Zimmermann *et al.*, 2009; Klein, 2011; Paterson *et al.*, 2011). Of these, four are under complete control (*Cereus jamacaru* D. C., *Cylindropuntia fulgida* Engelman, *C. leptocaulis* Knuth, and *Harrisia martinii* Labour) and eight are under substantial control (*Cylindropuntia imbricata* Haworth, *Harrisia balansae* (K.Schum.) N. P. Taylor & Zappi, *Opuntia aurantiaca* Lindley, *O. engelmannii* Salm-Dyck ex. Engelm., *O. ficus-indica* (L.) Mill., *O. monacantha* Haworth, *O. salmiana* J. Parm. ex Pfeiff, and *O. stricta* Haworth) (Klein, 2011).

From 1920 to 1935, the Australian government funded the importation of 52 different potential cactus biocontrol agents for host specificity testing; of which 12 were released (Raghu and Walton, 2007). These included multiple *Dactylopius* species, two *Chelinidea* spp. (Hemiptera: Coreidae), an *Olycella* sp. (Lepidoptera: Pyralidae), and *Tetranychus opuntiae* (Acari: Tetranychidae) (Raghu and Walton, 2007). Additionally, *Cactoblastis cactorum* (Lepidoptera: Pyralidae) was released in 1925 and established successfully (Mann, 1970); destroying hundreds of hectares of *Opuntia*-infested areas within two years (Raghu and Walton, 2007). 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). Twenty-seven species are listed in the Weeds of National Significance (WoNS) database (Table 1.2) (Sheehan and Potter, 2017). Of the naturalised species, *Corynopuntia* sp., *O. dejecta*, and *O. ficus-indica* are not considered invasive (Sheehan and Potter, 2017). The two most prominent biological control agents in use are *C. cactorum* and four *Dactylopius* species; namely *D. austrinus*, *D. ceylonicus*, *D. opuntiae*, and four *D. tomentosus* lineages (Sheehan and Potter, 2017).

The absence of indigenous Cactaceae and its relatives outside the New World allows for the use of less host-specific (oligophagous) agents in invaded areas, as there is a lower risk of non-target attack (Zimmermann and Granata, 2002). This also means that multiple target weeds can be attacked by oligophagous agents (Paterson *et al.*, 2019).

**Table 1.1:** National Environmental Management: Biodiversity Act (NEM:BA) Category 1 and 2 invasive cactus species in South Africa. Category 1a and 1b weeds require compulsory control, and need to be removed and destroyed. Category 2 weeds are regulated by area, and require demarcation permits (none are, however, allowed in riparian zones) (Department of Environmental Affairs, 2016).

Species	Common name	Category
<i>Austrocylindropuntia cylindrica</i> (Juss. ex Lam.) Backeberg.	Cane cactus	1a
<i>Austrocylindropuntia subulata</i> (Muehlenpf.) Backeb. subsp. exaltata (A. Berger) D. R. Hunt	Long spine cactus	1b
<i>Cereus hexagonus</i> (L.) Mill.	Queen of the night	1b
<i>Cereus hildmannianus</i> K. Schum.	Queen of the night	1b
<i>Cereus jamacaru</i> D. C.	Queen of the night	1b
<i>Cylindropuntia fulgida</i> (Engelm.) F. M. Knuth var. <i>fulgida</i>	Chain-fruit cholla	1b
<i>Cylindropuntia fulgida</i> (Engelm.) F. M. Knuth var. <i>mamillata</i> (Schott ex Engelm.)	Boxing glove cactus	1b
<i>Cylindropuntia imbricata</i> (Haw.) F. M. Knuth	Imbricate cactus	1b
<i>Cylindropuntia leptocaulis</i> (D. C.) F. M. Knuth	Pencil cactus	1b
<i>Cylindropuntia pallida</i> (Rose) F. M. Knuth	Pink-flowered sheathed cholla	1a
<i>Cylindropuntia spinosior</i> (Engelm.) F. M. Knuth	Cane cholla	1a
<i>Harrisia balansae</i> (K. Schum.) N. P. Taylor & Zappi	Strangler prickly apple	1a
<i>Harrisia martinii</i> (Labour.) Britton	Moon cactus	1b
<i>Harrisia pomanensis</i> (F. A. C. Weber) Britton & Rose	Midnight lady	1a
<i>Harrisia tortuosa</i> (J. Forbes ex Otto & A. Dietr.) Britton & Rose	Spiny snake cactus	1b
<i>Hylocereus undatus</i> (Haw.) Britton & Rose	Night-blooming cereus	2
<i>Myrtillocactus geometrizans</i> (Mart.) Console	Bilberry cactus	1a
<i>Opuntia aurantiaca</i> Lindl.	Jointed cactus	1b
<i>Opuntia elata</i> Link & Otto ex Salm-Dyck	Orange tuna	1b
<i>Opuntia engelmannii</i> Salm-Dyck ex Engelm.	Small round-leaved prickly pear	1b
<i>Opuntia ficus-indica</i> (L.) Mill.	Sweet prickly pear	1b
<i>Opuntia humifusa</i> (Raf.) Raf	Creeping prickly pear	1b
<i>Opuntia leucotricha</i> D. C.	Aaron's-beard prickly pear	1b
<i>Opuntia microdasys</i> (Lehm.) Pfeiff	Yellow bunny-ears	1b
<i>Opuntia monacantha</i> Haw	Drooping prickly pear	1b
<i>Opuntia pubescens</i> J. C. Wendl. ex Pfeiff.	Velvet bur cactus	1a
<i>Opuntia robusta</i> H. L. Wendl. ex Pfeiff	Blue-leaf cactus	1a
<i>Opuntia salmiana</i> J. Parm. ex Pfeiff	Bur cactus	1a
<i>Opuntia spinulifera</i> Salm-Dyck	Saucepan cactus	1b
<i>Opuntia stricta</i> (Haw.) Haw. var. <i>stricta</i> and var. <i>dillenii</i> (Ker Gawl.) L. D. Benson	Pest pear of Australia	1b
<i>Opuntia tomentosa</i> Salm-Dyck	Velvet opuntia	1b
<i>Peniocereus serpentinus</i> (Lag. & Rodr.) N. P. Taylor	Serpent cactus	1b
<i>Pereskia aculeata</i> Mill.	Barbados gooseberry	1b
<i>Tephrocactus articulatus</i> (Pfeiff.) Backeb.	Pine cone cactus	1a

**Table 1.2:** List of the Cactaceae in the Australian Weeds of National Significance (WoNS) database (Sheehan and Potter, 2017).

Species	Common name
<i>Austrocylindropuntia cylindrica</i>	Cane cactus
<i>Austrocylindropuntia subulata</i>	Eves needle cactus
<i>Cylindropuntia fulgida</i> var. <i>mamillata</i>	Coral cactus
<i>Cylindropuntia imbricata</i>	Devils rope
<i>Cylindropuntia kleiniae</i>	Klein's cholla
<i>Cylindropuntia leptocaulis</i>	Pencil cactus
<i>Cylindropuntia pallida</i>	White-spined Hudson pear
<i>Cylindropuntia prolifera</i>	Jumping cholla
<i>Cylindropuntia spinosior</i>	Snake cactus
<i>Cylindropuntia tunicata</i>	Brown-spined Hudson pear
<i>Opuntia aurantiaca</i>	Tiger pear
<i>Opuntia elata</i>	Riverina pear
<i>Opuntia elatior</i>	Red-flower prickly pear
<i>Opuntia engelmannii</i>	Engelmann's prickly pear
<i>Opuntia humifusa</i>	
<i>Opuntia leucotricha</i>	
<i>Opuntia microdasys</i>	Bunny ears
<i>Opuntia</i> sp. aff. <i>microdasys</i>	
<i>Opuntia monacantha</i>	Drooping tree pear
<i>Opuntia</i> aff. <i>polyacantha</i>	
<i>Opuntia puberula</i>	
<i>Opuntia robusta</i>	
<i>Opuntia schickendantzii</i>	
<i>Opuntia streptacantha</i>	Westwood pear
<i>Opuntia stricta</i> var. <i>stricta</i> and var. <i>dillenii</i>	
<i>Opuntia sulphurea</i>	
<i>Opuntia tomentosa</i>	Velvet tree pear

### *Opuntia* Mill

The *Opuntia* genus is the most widespread of all the cacti, comprising approximately 180 species and numerous hybrids (Anderson, 2001; Griffith, 2003). *Opuntia ficus-indica* L. ('sweet prickly pear') was likely one of the first cactus species introduced to South Africa in the early 1700s, and became widespread throughout the Karoo and Eastern Cape by the 1800s (Annecke and Moran, 1978; Dean and Milton, 2000). Numerous spineless varieties were bred in the early 1900s by the horticulturist Luther Burbank, and exported around the world as a source of food and fodder for livestock, and for its fruit (Anderson and Olsen, 2015). The first record of an *Opuntia* species in Australia was in 1788, when the British brought both *O. monacantha* and *Dactylopius* species into the country to establish a red dye industry (Mann, 1970). Many more *Opuntia* species were introduced over the next few decades, and proceeded to cause major environmental damage and economic losses to the agricultural sectors in South Africa and Australia due to the growth of dense, impenetrable thickets on farmlands (Dean and Milton, 2000). Novoa *et al.* (2019)

showed that spineless varieties grown from seeds (i.e. sexual reproduction) reverted back to their spiny form, which explains their high incidence in invaded regions. Mechanical and chemical control methods were initially implemented to control the weed, particularly the use of arsenical herbicides such as injectable monosodium methanearsonate (MSMA) (Annecke and Moran, 1978; Zimmermann and Moran, 1991). *Opuntia stricta* Haworth (var. *stricta*) ('Australian pest pear') was first recorded in Australia in 1839, and spread at a rate of 320 000 ha per year (at densities of about 250 000 kg per ha!) by the 1920s (Raghu and Walton, 2007). *Dactylopius ceylonicus* was initially released to combat the weed in 1903, but did not establish because this insect is host-specific to *O. monacantha*. The later release of the *D. opuntiae* 'stricta' lineage in 1921 caused substantial damage (Winston *et al.*, 2014). Following the success of the biological control of *Opuntia stricta* Haworth (both subspecies, nameley var. *stricta* and var. *dillenii*) and *O. ficus-indica* L. in Australia, four insects were successfully released in South Africa as an additional means to control the weed (Annecke and Moran, 1978; Zimmermann and Moran, 1991). These were the cladode boring moth *C. cactorum* Berg (Lepidoptera: Pyralidae) released in 1933, the cladode sucking cochineal bug *D. opuntiae* ('ficus' lineage) Cockerell (Hemiptera: Dactylopiidae) released in 1938, and the stem boring beetles *Lagocheirus funestus* Thompson (Coleoptera: Cerambycidae) released in 1943 and *Metamasius spinolae* Gyllenhal (Coleoptera: Curculionidae) released in 1948 (Klein, 2011). Of these, *C. cactorum*, and particularly *D. opuntiae*, have been the most effective agents - to the point that the agents themselves were seen as a threat to commercial *O. ficus-indica* plantations (Zimmermann and Moran, 1991). This is currently the case in Brazil, Israel, and the Mediterranean area, where *D. opuntiae* is threatening and/or causing substantial losses to the cactus fodder and fruit industry (Spodek *et al.*, 2014; Torres and Giorgi, 2018; Mazzeo *et al.*, 2019). Conflicts of interest are common in the control of invasive Cactaceae (Novoa *et al.*, 2016), and arise when an invasive species holds economic value in its introduced range. For example, *Opuntia ficus-indica* L. (Mill.) provides a source of income for many rural communities in South Africa (Brutsch and Zimmermann, 1993; Beinart and Wotshela, 2011). The fruit is harvested and sold, and is also used to make beer, jam, and traditional medicines (Brutsch and Zimmermann, 1993; Shackleton *et al.*, 2011). Rural communities in the Eastern Cape Province in particular rely heavily on the fruit for an income, and the local people are generally unaware of its invasive status (Shackleton *et al.*, 2007). Greater communication between these informal traders and the municipality is necessary, where legislation needs to be implemented that achieves a balance between limiting the negative

effects of the invader on the environment, and promoting economic growth (Shackleton *et al.*, 2011). *Opuntia stricta* became particularly problematic in the Kruger National Park in South Africa; where elephants, baboons, and rivers aid in its dispersal (Hoffmann *et al.*, 1998b). *Cactoblastis cactorum* was released in the park in 1988 to control the weed (Hoffmann *et al.*, 1998a), followed by *D. opuntiae* ('stricta' lineage) in 1997 (Klein, 2011). Both these insects established successfully, where a biomass reduction of over 90%, and an almost complete end to fruit production, has been recorded (Paterson *et al.*, 2011). The weed is currently considered to be under substantial control (Klein, 2011).

*Opuntia aurantiaca* Lindley, commonly referred to as 'jointed cactus', is indigenous to Argentina and Uruguay, and is thought to be a hybrid of *O. discolor* Britton & Rose and *O. salmiana* J. Parm ex. Pfeiff (Moran and Zimmermann, 1991a). It is now accepted that *O. aurantiaca* is native to east Argentina and Uruguay (De Lotto, 1974; Gunn, 1979; Zimmermann and van de Venter, 1981). The weed was introduced to South Africa as an ornamental plant in 1843 from collections in the United Kingdom (Zimmermann and van de Venter, 1981), and later became problematic in grazing areas in the Karoo region and the Eastern Cape Province (Moran and Zimmermann, 1991a). The first biocontrol agent to be released on this weed in Australia and South Africa was *D. austrinus* De Lotto (Hemiptera: Dactylopiidae) in 1933 and 1935, respectively (Moran and Zimmermann, 1991a; Winston *et al.*, 2014). Only *D. austrinus* and *C. cactorum* established successfully on the weed (Klein, 2011), while the other three attempted agents failed (*Mimorista pulchellalis* Dyar (Lepidoptera: Crambidae), *Nanaia* spp. (Lepidoptera: Pyralidae), and *Zophodia tapiacola* Dyar (Lepidoptera: Pyralidae)).

### ***Cylindropuntia* (Engelmann) (Knuth)**

The *Cylindropuntia* genus (commonly known as 'chollas') is a monophyletic clade that falls within the Opuntioideae subfamily (Anderson, 2001). It was originally classified as a subgenus of *Opuntia* in 1856, but was later placed in its own genus in 1935 (Anderson, 2001). The genus comprises 33 species, and, as with many of the *Opuntia* species, undergoes natural hybridisation (Anderson, 2001). *Cylindropuntia imbricata* was first introduced to South Africa in the early 1900s (Moran and Zimmermann, 1991a). The weed was successfully controlled in Australia by *D. tomentosus* after its release in the country in 1925 (Dodd, 1940). South Africa followed suit and released the insect in 1970 (Moran and Zimmermann, 1991a). The only other biocontrol agent associated with the weed, *Metamasius spinolae*, did not establish successfully (Klein,

2011).

*Cylindropuntia fulgida* (Engelm. Knuth) (chain fruit cholla) is a tree-like cactus that comes in two varieties; namely var. *fulgida* and var. *mamillata* (Anderson, 2001). *Cylindropuntia fulgida* var. *fulgida* was introduced to South Africa in the 1940s as an ornamental plant (de Beer, 1986), and has become particularly damaging to pastoral lands in the Northern Cape and Limpopo Provinces (Moran and Zimmermann, 1991a; Paterson *et al.*, 2011). The cactus was initially misidentified as *C. rosea*, and as a result, biocontrol efforts were largely unsuccessful due to a less effective *D. tomentosus* lineage being released (Paterson *et al.*, 2011). Following the correct identification of the weed, various *D. tomentosus* lineages were collected in Mexico; namely from the two *C. fulgida* varieties and from *C. cholla* (F.A.C. Weber) (Knuth) (Paterson *et al.*, 2011). It was found that the *C. cholla* lineage was the most damaging, and hence became the agent of choice for this cactus variety (Mathenge *et al.*, 2009b). Other category 1a and 1b species in this genus in South Africa are *C. pallida*, *C. spinosior*, and *C. leptocaulis* (Table 1.1). In addition to these, Australia has listed *C. kleiniae*, *C. prolifera*, and *C. tunicata* in their WoNS database (Table 1.2).

## 1.4 The Dactylopiidae

The Dactylopiidae family (Hemiptera: Sternorrhyncha: Coccoidea) comprises 11 species and a variety of intraspecific lineages in the *Dactylopius* genus; which are all indigenous to the Americas (Rodríguez *et al.*, 2001; Van Dam and May, 2012) (Table 1.3).

**Table 1.3:** A list of the 11 recorded *Dactylopius* species and intraspecific lineages

Species
<i>Dactylopius austrinus</i> De Lotto
<i>Dactylopius bassi</i> Targioni Tozzetti
<i>Dactylopius ceylonicus</i> Green
<i>Dactylopius coccus</i> Costa
<i>Dactylopius confertus</i> De Lotto
<i>Dactylopius confusus</i> Cockerell
<i>Dactylopius gracilipilus</i> Van Dam & May
<i>Dactylopius opuntiae</i> Cockerell
Lineages: ‘ficus’, and ‘stricta’
<i>Dactylopius tomentosus</i> Lamarck
Lineages: ‘bigelovii’, ‘californica var. parkeri’, ‘cholla’, ‘cylindropuntia sp.’, ‘echinocarpa x acanthocarpa’, ‘imbricata’
<i>Dactylopius salmianus</i> De Lotto
<i>Dactylopius zimmermannii</i> De Lotto

Four of these are currently used as biological control agents in South Africa and Australia; namely *D. austrinus* De Lotto, *D. ceylonicus* Green, *D. opuntiae* Cockerell, and *D. tomentosus* Lamarck (De Lotto, 1974; Pérez-Guerra and Kosztarab, 1992; Sheehan and Potter, 2017). All species produce carminic acid, a deep red anthraquinone, as a defensive compound (Eisner *et al.*, 1980; Pérez-Guerra and Kosztarab, 1992). The Aztecs and Incas domesticated *D. coccus* as a dye for wool, cotton, paint, and cosmetics (Nobel, 2002). The dye has been used for over two thousand years, with the first historical reports from Peru dating back to the early 1500s (Nobel, 2002). The Spanish conquistadors initiated large-scale cultivation of the insects in Oaxaca, Mexico, and began exporting them to Europe by the year 1526 (Meyer and Beezley, 2000). The dye became the third highest source of income to the Spanish Empire after gold and silver (Humboldt, 1966; Chávez-Moreno *et al.*, 2009).

### 1.4.1 Identification issues

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. 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 Viguera, 2006). Considering the high species diversity of the Cactaceae, the probability of there being many more *Dactylopius* species, cryptic species, and lineages in the native range is very high. Using molecular tools to assist in the taxonomic organisation of this group is therefore very valuable. Different *Dactylopius* species and lineages display different levels of damage on target 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.

### 1.4.2 'Biotypes'

Species within the Dactylopiidae consist of genetically and/or behaviourally-distinct groups, which have been referred to as 'biotypes' in the literature (Githure *et al.*, 1999; Volchansky *et al.*, 1999; Hoffmann, 2004; Mathenge *et al.*, 2009a, 2010a,b; Jones *et al.*, 2015; Mathenge *et al.*, 2015). Mathenge *et al.* (2010b), for example, suggest that "*biotypes of Dactylopius species may not have developed sufficient behavioural, physiological or genetic changes during their prolonged association with their respective hosts*

to cause reproductive isolation and should be regarded as biotypes rather than as races.”, while a more recent paper by Mathenge *et al.* (2015) concludes that “...currently described *Dactylopius* species may be species complexes composed of biotypes, host races, cryptic, or sibling species”. Volchansky *et al.* (1999) use the terms ‘biotype’, ‘strain’, and ‘host race’ interchangeably regarding the two distinct ‘ficus’ and ‘stricta’ *D. opuntiae* lineages. A review by Miller and Kosztarab (1979) refers to ‘demes’ of scale insect populations that arise due to differences in host-plant defenses, such as with *Dynaspidiotus californica* Coleman (Hemiptera: Diaspididae) on *Pinus ponderosa* Douglas (Edmunds and Alstad, 1978).

Numerous definitions exist for the term ‘biotype’, synonymous with the species concept debate (Häuser, 1987). Eastop (1973) states that a “‘*Biotype*’ is a taxonomic concept mostly used by non-taxonomists and has been defined as consisting of all individuals of equal genotype. Biotypes are recognised by a biological function rather than by morphological characters. In practice a biotype contains those individuals performing whatever biological feat interests the observer and thus may contain one or more races or strains.” Maxwell and Jennings (1980) define the term as “an individual or population that is distinguished from the rest of its species by criteria other than morphology, for example, a difference in parasite ability”, while Diehl and Bush (1984) define it as “Entomophagous or phytophagous parasites or parasitoids distinguished by survival and development on a particular host or by host preference for feeding, oviposition, or both. Other insect biotypes differ in diurnal or seasonal activity patterns, size, shape, color, insecticide resistance, migration and dispersal tendencies, pheromone differences, or disease vector capacities.”. A review by Downie (2010) labels this term as a pseudo-taxonomic classification that is too simplistic and misleading. Similarly, Claridge and Den Hollander (1983) suggested that the term ‘biotype’ is a convenient one to cover the lack of understanding of particular plant-insect interactions, that it is confusingly applied to both individuals and populations within a species, and that it incorrectly implies that these genetic groups are discrete entities. The concept of biotypes was first mentioned by Printz (1937) in describing differences in the virulence of *Phylloxera* populations, and again by Painter (1941) for the description of different geographic populations of Hessian flies (*Mayetiola destructor* Say (Diptera: Cecidomyiidae)) that differed in their virulence to wheat. Downie (2010) discusses a number of other agricultural insect pests that have been designated into multiple biotypic groups, including the greenbug *Schizaphis graminum*

Rondani (Hemiptera: Aphididae) (Porter *et al.*, 1997; Burd and Porter, 2006), the rice brown planthopper *Nilaparvata lugens* Stal (Hemiptera: Delphacidae) (Naeemullah *et al.*, 2009), grape phylloxera (*Daktulosphaira vitifoliae* Fitch (Hemiptera: Phylloxeridae)) (Stevenson, 1970), and the whitefly (*Bemisia tabaci* Gennadius (Hemiptera: Aleyrodidae)) (Perring, 2001). Numerous definitions and terms have been used in conjunction, and synonymously with, that of ‘biotype’, adding to the confusion. Some of these follow below along with their given definitions taken from various sources in the literature:

**Biological strain:** *“An isolate or group of isolates that can be distinguished from other isolates of the same genus and species by phenotypic characteristics or genotypic characteristics or both.”* (Tenover *et al.*, 1995)

*“A natural or artificial mating group detected by difference in parasitic adaptation.”* (Darlington and Mather, 1952)

**Clade:** *“Branch of a phylogenetic tree containing the set of all organisms descended from a particular common ancestor which is not an ancestor of any non-member of the group.”* (Lawrence, 2011)

*“In cladistics, a lineage branch that results from splitting in an earlier lineage. A split produces two distinct new taxa, each of which is represented as a branch in a phylogenetic diagram. The term is derived from the Greek klados meaning ‘twig’ or ‘branch’.”* (Allaby, 1992)

**Cryptic species:** *“Two or more species that have been classified as a single nominal species because they are at least superficially morphologically indistinguishable.”* (Bickford *et al.*, 2007)

**Deme:** *“A local population unit of a species within which breeding is completely random.”* (Lawrence, 2011)

*“A spatially discrete, interbreeding group of organisms with definable genetic or cytological characters (i.e. a subpopulation of a species). There is very restricted genetic exchange, if any, with other demes, although demes are usually contiguous with one another, unlike subspecies or races, which are often isolated by*

*some geographical or habitat barrier.*" (Allaby, 1992)

**Ecotype:** *"The ecological unit arising as a result of the genotypical response of an ecospecies to a particular habitat."* (Turrill, 1946)

*"A genetically unique population that is adapted to its local environment."* (Turesson, 1922)

*"A subspecific form within a true species, resulting from selection within a particular habitat and therefore adapted genetically to that habitat, but which can interbreed with other members of the species."* (Lawrence, 2011)

**Host race:** *"1) uses different host taxa in the wild, 2) consists of individuals that exhibit 'host fidelity', 3) coexists in sympatry [with other host races] in at least part of their ranges, 4) is genetically differentiated at more than one locus, 5) is spatially and temporally replicable, 6) displays a correlation between host choice and mate choice, 7) undergoes actual gene flow (hybridization and backcrossing), 8) has higher fitness on natal than alternative hosts and 9) [in some cases] produces hybrids that are less fit than parental forms."* (Drès and Mallet, 2002)

*"A population of a species that is partially reproductively isolated from other conspecific populations as a direct consequence of adaptation to a specific host. The basis of isolation may involve genetically based differences in host preference or many other factors such as allochronic [= occurring in different segments in geologic time] barriers that arise as a direct result of phenological differences among hosts."* (Diehl and Bush, 1984)

**Lineage:** *"An ancestral-descendant sequence of interbreeding populations."* (Simpson, 1951)

*"An evolutionary unit that includes an ancestral population, all of its descendants, and only its descendants. Also called a monophyletic group or a clade."* (Freeman, 2005)

*"... ultimately extends back through the various taxonomic levels, from the species to the genus, from the genus to the family, from the family to the order, etc."* (Allaby, 1992)

**Morph:** *"Discrete phenotypic variants that segregate within a population."* (Mayr, 1970)

**Race:** “... where the individuals of a species can be divided into groups, usually isolated to some extent by food preferences, occurring in the same locality and showing definite differences in biology, but with corresponding structural differences, either few or inconsistent, or completely absent.” (Thorpe, 1930)

“A genetically, and as a rule geographically, distinct mating group within a species.” (Darlington and Mather, 1952)

“A group of individuals within a species which forms a permanent and genetically distinguishable variety.” (Lawrence, 2011)

“A population that has different characteristics from another population of the same species, whether or not there are significant genetic differences between them.” (Freeman, 2005)

**Sibling species:** “Morphologically similar or identical populations which are reproductively isolated.” (Mayr, 1963)

“True species which do not interbreed but are difficult to separate on morphological grounds alone.” (Lawrence, 2011)

**Variety:** “A sub-division of a species owing its uniformity either to genetic isolation in nature, or to artificial propagation in cultivation (where a variety is often a clone).” (Darlington and Mather, 1952)

“Taxonomic group below the species level.” (Lawrence, 2011)

In agreement with the view presented by Diehl and Bush (1984) and Downie (2010) regarding the superfluous nature of the term ‘biotype’, this thesis instead makes use of the term ‘lineage’ relative to the particular genetic markers used to differentiate between clades and genotypic clusters.

### 1.4.3 Species summary

The Dactylopiidae occur across the Americas; where five species occur in North America, and five in South America (Rodríguez *et al.*, 2001). *Dactylopius coccus* has a disjoint distribution, being present in both these regions (referred to as an ‘amphitropical’ distribution) (Rodríguez *et al.*, 2001; Van Dam *et al.*,

2015). *Dactylopius bassi* was collected in Mexico and initially placed in the *Coccus* genus in 1867, but was later reclassified as a member of the Dactylopiidae by Ben-Dov and Marotta (2001). The type material for this species has since been lost and it cannot at present be distinguished from the other *Dactylopius* species; making its classification somewhat dubious. See Section 6.2 for a description of the life-cycle of the Dactylopiidae.

### ***Dactylopius coccus* Costa**

This is a domesticated species that has undergone artificial selection for well over a thousand years (Gon and Price, 1984; Chávez-Moreno *et al.*, 2009), and is the only *Dactylopius* species with 16 chromosomes; the other recorded species have 10 (Gavrilov, 2007). It was used as a source of red dye by the Aztecs since at least the tenth century, and was cultivated on a mass scale by the Spanish during their conquest of Mexico in the 1500s (Mann, 1969; Pérez-Guerra and Kosztarab, 1992; Greenfield, 2005; Chávez-Moreno *et al.*, 2009). *Dactylopius coccus* produces the largest amount of carminic acid of all the *Dactylopius* species (Chávez-Moreno *et al.*, 2009). Its origin is currently unknown, and is disputed in the literature. Rodríguez *et al.* (2001) suggest that it originated in South America, and was introduced to North America through trade during pre-Columbian times. Contrastingly, Campana *et al.* (2015) suggest that the species derives from at least two population sources; namely in Mexico and Peru. The original source population is unknown, although Portillo (2007) suggests that it originated in North America due to the presence of all its host plants and natural enemies there.

The host plant most commonly used to rear this insect is *Opuntia ficus-indica* (Donkin, 1977; Portillo and Viguera, 2006); which is a major cause of the spread of this weed around the world. *Dactylopius coccus* was released as a biological control agent in Australia in 1926, but it did not establish successfully (Dodd, 1940; Winston *et al.*, 2014). Other recorded host plants include *Nopalea cochenillifera* L. (Mill.), *O. atropes* Rose, *O. crassa* Haw., *O. fulginosa* Griffiths, *O. hyptiacantha* F. A. C. Weber, *O. jaliscana* Bravo, *O. pilifera* F. A. C. Weber, *O. robusta* Wendl. ex Pfeiff., *O. steptacantha* Lem., *O. tomentosa* Salm-Dyck and *O. undulata* Griffiths (Mann, 1969; Chávez-Moreno *et al.*, 2011).

## NORTH AMERICA

***Dactylopius confusus* Cockerell**

This species covers a wide distribution from Canada to Texas, as well as California, Mexico, and Florida (Gilreath and Smith, 1987). Recorded host plants are *Cylindropuntia imbricata* (Haw.) Knuth, *Cylindropuntia kleiniae* (D. C.) Knuth, *Cylindropuntia leptocaulis* (D. C.) Knuth, *Cylindropuntia tunicata* (Lehm.) Knuth, *Grusonia grahamii* (Engelm.) H. Rob., *Opuntia ficus-indica* (L.) Mill., *Opuntia fuliginosa* Griffiths, *Opuntia hyptiacantha* F. A. C. Weber, *Opuntia jaliscana* Bravo, *Opuntia phaeacantha* Engelm., *Opuntia pubescens* H. L. Wendl. ex Pfeiff., *Opuntia spinulifera* Salm-Dyck, and *Opuntia streptacantha* Lem. (Mann, 1969; Chávez-Moreno *et al.*, 2011). The insect was introduced to Australia in 1915 and 1926, South Africa in 1832, and India in 1836 and 1838 as a biological control agent of *Opuntia monacantha* (Willd.) Haw, but it did not successfully establish in any of these countries (Dodd, 1940; Winston *et al.*, 2014).

***Dactylopius gracilipilus* van Dam & May**

*Dactylopius gracilipilus* was described as a new species in 2012 by Van Dam and May (2012). The insects were found on *Corynopuntia schottii* Engelm. in the Chihuahuan Desert, Texas. This species is host-specific to the Opuntioid genus *Corynopuntia* Knuth, and is very morphologically similar to *D. tomentosus* (Van Dam and May, 2012).

***Dactylopius opuntiae* Cockerell**

*Dactylopius opuntiae* is indigenous to Mexico, and the southern regions of Texas, New Mexico, and Arizona (Mann, 1969). Its recorded host plants are *C. imbricata*, *C. tunicata*, *N. cochenillifera*, *N. karwinskiana*, *O. atropes*, *O. fuliginosa*, *O. hyptiacantha*, *O. jaliscana*, *O. joconostle*, *O. leucotricha*, *O. macdougaliana*, *O. megacantha*, *O. phaeacantha*, and *O. robusta*. (Mann, 1969; Chávez-Moreno *et al.*, 2011). Numerous consignments of the insect were introduced to Australia between 1921 and 1935 to control *O. stricta* Haw., *O. streptacantha* Lem., and *O. tomentosa* Salm-Dyck (Dodd, 1940; Winston *et al.*, 2014). Other releases took place in Madagascar in 1923, Sri Lanka in 1925, India in 1926, Mauritius in 1928, Indonesia in 1935, Hawaii in 1949 and 1950, the Federation of St Kitts

and Nevis in 1957, and Kenya from 1958 onwards (Winston *et al.*, 2014). The insect was introduced to South Africa from an Australian stock (originally collected in Mexico and Arizona) in 1938, with further releases taking place in the 1980s from Australian stocks collected in Texas and Arizona, to control *O. ficus-indica*, *O. stricta*, and *O. engelmannii* (Volchansky *et al.*, 1999; Winston *et al.*, 2014). It established on *O. ficus-indica* and *O. engelmannii*, but not on *O. stricta*, despite successes on the plants in Australia (Volchansky *et al.*, 1999). It was initially thought that Australian and South African *O. stricta* host plants were genetically distinct, but the confusion arose due to the fact that the Australians had unknowingly imported multiple lineages of *D. opuntiae* from the Americas during the 1920s and 1930s (Dodd, 1940; Volchansky *et al.*, 1999). Insects were collected from and reared on different species of host plants prior to transportation (most commonly *O. engelmannii* var *lindheimeri* Engelm.) (Dodd, 1940). It is believed that the lineage originally sent to South Africa was collected from *Opuntia streptacantha* Lem. in Mexico, which is more closely related to *O. ficus-indica* than to *O. stricta* (Volchansky *et al.*, 1999). This was the ‘ficus’ lineage. *Dactylopius opuntiae* ‘stricta’ was collected from Tamworth in Australia, and released in South Africa in 1997 to control *Opuntia stricta*, and again in 2000 to combat *O. humifusa* Raf., with successful establishment on the weeds (Hoffmann, 2003; Rule and Hoffmann, 2018). Githure *et al.* (1999) showed that the ‘stricta’ lineage was unable to develop on *O. ficus-indica*, with the exception of one spineless cultivar (the American Giant). Development and survival on this cultivar, was, however lower than on the preferred *O. stricta* host plant.

The presence of both the ‘ficus’ and ‘stricta’ lineages in South Africa has led to cases of hybridisation, which affects the host-specificity of progeny (Hoffmann *et al.*, 2002; Hoffmann, 2003). Hoffmann *et al.* (2002) found that F1 generation hybrids lost their host-specificity and could develop equally as well on both *O. ficus-indica* and *O. stricta*, while crossing hybrids with true-bred individuals in the F2 generation yielded progeny that were either all true-bred (host-specific), hybrids (not host-specific), or a mixture of the two. This inheritance dynamic is attributed to the unusual lecanoid genetic system of many scale insects, where males only pass on their maternal genes to progeny (see Section 6.1 for further details). The study concluded that although non-host specific hybrids could be beneficial to the biological control of both cactus weeds, the production of host-specific F2 progeny could result in agents on an incompatible plant. It is thus recommended that the two insect lineages should be mass-reared and

released separately.

### ***Dactylopius tomentosus* Lamarck**

This species is native to Arizona, New Mexico, Mexico, and Texas (Rodríguez *et al.*, 2001). Recorded host plants are *C. acanthocarpa* Knuth, *C. bigelovii* (Engelm.) Knuth, *C. leptocaulis* Knuth, *C. tunicata* (Lehm.) Knuth, *Nopalea karwinskiana* Salm-Dyck., *C. atropes* = *velutina* Rose/F. A. C. Weber, and *O. ficus-indica* (L.) Mill. (Mathenge *et al.*, 2009a; Chávez-Moreno *et al.*, 2011). Two *Dactylopius tomentosus* lineages have been released in Australia, namely ‘imbricata’ (released in 1925 from a population collected in Texas in the USA (Winston *et al.*, 2014)) and ‘cholla’ (approved for release in 2016 (Sheehan and Potter, 2017)), to combat *C. imbricata* and *C. fulgida*, respectively (Jones *et al.*, 2016b). The ‘cholla’ lineage was obtained in 2011 from Musina in the Limpopo Province in South Africa; which was originally collected in Loretto and La Paz (Baja California Sur, Mexico) (Day *et al.*, 2015). *Cylindropuntia imbricata* is the only *Cylindropuntia* species to have been successfully controlled in Australia to date, making the search for additional *D. tomentosus* lineages to combat the other seven species a priority (Jones *et al.*, 2015, 2016b,a). Jones *et al.* (2016a) screened an additional four populations collected from different *Cylindropuntia* host plants in the southern USA. These were namely ‘leptocaulis’, ‘acanthocarpa’, ‘echinocarpa x acanthocarpa’ (a suspected hybrid of *C. echinocarpa* and *C. acanthocarpa*), and ‘cylindropuntia sp.’ from an undetermined *Cylindropuntia* host plant. Three lineages were approved for release in 2017; namely the ‘bigelovii’ lineage for the control of *Cylindropuntia spinosior*, the ‘cylindropuntia sp.’ lineage for *C. kleini* and *C. leptocaulis*, and the ‘acanthocarpa x echinocarpa’ lineage for *C. tunicata* (Sheehan and Potter, 2017).

There are two predominant lineages of *D. tomentosus* in South Africa, namely ‘cholla’ (effective only on *Cylindropuntia fulgida* var. *fulgida* and var. *mammilata*) and ‘imbricata’ (effective only on *Cylindropuntia imbricata*) (Mathenge *et al.*, 2010b). The ‘imbricata’ lineage was first introduced to South Africa in 1970 from stocks collected in Texas, USA (via Australia) (Moran and Zimmermann, 1991b). The ‘cholla’ lineage was subsequently released after it was found that the ‘imbricata’ lineage only minimally impacted *C. fulgida* var. *fulgida* (Moran and Zimmermann, 1991a). Establishment was successful, and resulted in moderate to high impact (Winston *et al.*, 2014). Two releases have also taken place in Zimbabwe in

2009 and 2011 for the control of *C. fulgida* var. *fulgida* and *C. fulgida* var. *mamillata* (Winston *et al.*, 2014).

Mathenge *et al.* (2009a) worked on five *D. tomentosus* source populations, and found that each group showed a host preference to particular *Cylindropuntia* species. This indicates that there are numerous potential lineages, and even possible cryptic species. Jones *et al.* (2015) also supported this multi-lineage finding. Hybridisation between the lineages occurs where *C. fulgida* var. *fulgida* and *C. imbricata* occur in sympatry, resulting in offspring that display differential host-specificities and fitness indices depending on the host plant associated with the maternal parent (Mathenge *et al.*, 2010b). Hybrids with a ‘cholla’ maternal genome displayed higher fitness on *C. fulgida*, while those with an ‘imbricata’ maternal genome displayed higher fitness on *C. imbricata*. Overall, the host-specificity of hybrids for their choice of either *C. fulgida* or *C. imbricata* was reduced. It is recommended that the two lineages be kept apart in the field until the long-term effects of hybridisation are better-understood (Paterson *et al.*, 2011).

## SOUTH AMERICA

### *Dactylopius austrinus* De Lotto

This insect is native to central, north, and western Argentina (De Lotto, 1974; Gunn, 1979; Zimmermann and van de Venter, 1981). Recorded host plants are *O. retrorsa* Speg., *O. discolor* Britton & Rose, *O. canina* Speg., *O. kiskaloro*, *O. palmadora* = *Tacinga palmadora*, *O. sulphurea* G. Don, and *O. salmiana* J. Parm. (Moran and Cabby, 1979). It was first released in Australia in 1933, and South Africa in 1935 to control *Opuntia aurantiaca* Lindley (jointed cactus) (Moran and Zimmermann, 1991a; Pérez-Guerra and Kosztarab, 1992). *Dactylopius austrinus* does not occur on *O. aurantiaca* in its native range, but rather on *O. utkilio* (= *elata*) Link & Otto and/or *O. sulphurea* G. Don (Moran and Cabby, 1979; Moran and Zimmermann, 1991a). This is because the host plant of the original *D. austrinus* stock brought to South Africa was collected from an unidentified host plant in the Catamarca province in Argentina. The insects were transferred to *O. aurantiaca* and then shipped to South Africa and Australia (Gunn, 1979).

***Dactylopius ceylonicus* Green**

This species is native to Argentina, Uruguay, and Brazil (Mann, 1969). Host plants in its native range include *O. quimilo* K. Schum., *O. bonaerensis* = *elata* Link & Otto, *O. discolor* Britton & Rose, *O. salmiana* J. Parm., and *O. monacantha* (Willd.) Haw. (Mann, 1969). Other recorded host plants include *C. imbricata*, *O. ficus-indica*, and *O. fuliginosa* (Chávez-Moreno *et al.*, 2011). The insect seldom occurs on *O. aurantiaca* Lindl., but can be cultivated on it (Mann, 1969). *Dactylopius ceylonicus* was the first insect released for biological control in South Africa in 1913 to combat *Opuntia monacantha* Haw. (Zimmermann *et al.*, 2004). It was released in Australia and Mauritius in 1914, in Tanzania in 1957, and Kenya in 1958 to control the same weed (Winston *et al.*, 2014).

***Dactylopius confertus* De Lotto, *Dactylopius salmianus* De Lotto, and *Dactylopius zimmermannii* De Lotto**

*Dactylopius confertus* is native to Argentina, where its recorded host plants include *Cleistocactus baumannii* Lem., *Echinopsis leucantha* Walp., *Cereus aethiops* Haw., *Denmoza rhodocantha* (Salm-Dyck) Britton & Rose, *Gymnocalycium monvillei* (Lem.) Britton & Rose, *Harrisia tortuosa* Britton & Rose, *Pilocereus* spp., and *Trichocereus candicans* Britton & Rose (Claps and de Haro, 2001). This wide range of host plants may indicate that this species contains a number of cryptic species, which could be revealed through genetic barcoding.

*Dactylopius salmianus* is native to Argentina, and host-specific to *Opuntia salmiana* J. Parm (Pérez-Guerra and Kosztarab, 1992). *Dactylopius zimmermannii* is native to Mendoza, Argentina (Pérez-Guerra and Kosztarab, 1992), and is host-specific to *Tephrocactus ovatus* Gill, *Maihueniopsis ovata* F. Ritter, *Cereus aethiops* Haw., *Maihuenia patagonica* Britton & Rose, and *Maihueniopsis darwinii* F. Ritter (Claps and de Haro, 2001).

## 1.5 DNA and molecular barcoding

### 1.5.1 History and application

Analogous to the Universal Product Codes (UPCs) found on commercial goods, genetic barcodes are created through the amplification and sequencing of short segments or gene regions of an organism's genome - most commonly the mitochondrial cytochrome oxidase I (COI) gene in animals, and the chloroplast *rbcL* (RuBisCo large subunit) and *matK* (maturase k) loci in plants (Hebert *et al.*, 2003a; CBOL, 2009; Floyd *et al.*, 2010). These unique signature sequences are used for identification and for phylogenetic analyses (Hebert *et al.*, 2003a). Mitochondrial genes are the preferred targets for barcoding initiatives due to their generally higher mutation rates compared to nuclear genes (Drake *et al.*, 1998; Ballard and Whitlock, 2004; Haag-Liautard *et al.*, 2008), their relatively low rates of recombination (Piganeau *et al.*, 2004), and their high abundance in cells (Waugh, 2007). The COI marker is particularly useful because it displays low rates of insertions and deletions, and despite being a relatively conserved region, it can be used to differentiate between species; particularly in insects (Hebert *et al.*, 2003a; Blaxter, 2004). Although the concept of species identification using molecular techniques began in the 1980s, particularly regarding microorganisms (Nanney, 1982; McAndrew and Majumdar, 1983; Anderson *et al.*, 1985), Arnot *et al.* (1993) first used the term 'DNA barcode' in a study aiming to barcode isolates of the *Plasmodium falciparum* Welch protozoan. The method gained popularity in 2003 (Hebert *et al.*, 2003a,b), and the Consortium for the Barcode of Life (CBOL) was founded the following year (CBOL, 2009). The Barcode of Life Database (BOLD) currently contains over 6 million genetic barcodes representing 215 000 animal, 69 000 plant, and 22 000 fungal species (BOLD, 2019). Also referred to as 'microgenomic identification', genetic barcoding relies on nucleotide diversity between organisms to group them taxonomically, or according to distinct haplotypes (Hebert *et al.*, 2003b). It is based on the premise that interspecific variation is greater than intraspecific variation, referred to as the 'barcoding gap' (Hebert *et al.*, 2003a). An arbitrary sequence divergence criterion suggested that species could be delineated at 3% for insects, and 2% and above for vertebrates (Hebert *et al.*, 2003a). DNA barcoding can serve as an efficient tool for taxonomists, and has a wide variety of cross-disciplinary applications (Savolainen *et al.*, 2005; Frézal and Leblois, 2008; Valentini *et al.*, 2008) (see Table B1). A major advantage of genetic barcoding is its applicability to various stages of an organism's life cycle; as the

genetic signature remains the same irrespective of differences in the individual's age (Floyd *et al.*, 2010).

### 1.5.2 Caveats

Despite the advantages, there are a number of criticisms of DNA barcoding, particularly with reference to its use as a standalone tool for the description of new species (Rubinoff *et al.*, 2006). The main arguments put forward include the the following:

1. Heteroplasmy and differential inheritance of mitochondrial genomes.
2. Maternal mitochondrial inheritance.
3. Differential rates of gene evolution.
4. Introgression of mitochondrial genes.
5. Incomplete lineage sorting.
6. The occurrence of nuclear mitochondrial DNA segments (numts).

#### **Heteroplasmy**

Heteroplasmy in the mitochondrial genome refers to the presence of more than one genotype within an organism, such that there is a ratio of 'mitotypes' (analogous to haplotypes); where one is often predominant over the others (Kmiec *et al.*, 2006). Mitotypes can come about due to differences in nucleotide length (length heteroplasmy) or composition of mtDNA (site heteroplasmy) due to the insertion or deletion of large fragments, or errors in the replication process (Barr *et al.*, 2005). New mitotypes can also be created following the exposure of mtDNA to oxygen metabolites, resulting in point mutations (Kmiec *et al.*, 2006). Mitochondria are the vestiges of ancestral bacterial endosymbionts (Gray *et al.*, 1999) and replicate independently of the nucleus (Meusel and Moritz, 1993). This unique system of genomic segregation and replication results in the differential transferal of mitochondrial genomes to daughter cells; facilitating heteroplasmy (Barr *et al.*, 2005).

#### **Maternal inheritance**

Although mitochondrial genomes are predominantly maternally inherited and rarely undergo recombination in animals, there are an increasing number of cases of biparental inheritance, or 'paternal leakage', in which some mitochondrial sequences are inherited from males (Gyllensten *et al.*, 1991; Zouros *et al.*, 1992; Skibinski

*et al.*, 1994; Hoarau *et al.*, 2002; Kvist *et al.*, 2003; Barr *et al.*, 2005; Xiong *et al.*, 2013; Ladoukakis and Zouros, 2017). This has been found in some insects, such as *Drosophila simulans* Sturtevant (Diptera: Drosophilidae) (Wolff *et al.*, 2013) and *Magicicada* spp. L. (Hemiptera: Cicadidae) (Fontaine *et al.*, 2007). Additionally, ‘doubly uniparental inheritance’ is another system that is common in bivalves, where male offspring inherit mtDNA from both parents while females inherit maternal mtDNA only (Zouros *et al.*, 1994). The paternal transmission of mtDNA is usually prevented through a number of mechanisms across species, such as the destruction of male mitochondria in the developing zygote, the total absence of mitochondria in sperm, and the prevention of male mitochondria entering the oocyte upon fertilisation (Rokas *et al.*, 2003). The process of paternal leakage, however, occurs more frequently in animals than was previously realised and can contribute to inconsistencies in the comparison of genetic barcodes. This is in addition to the growing understanding of the effects of mitochondrial bottlenecks during oogenesis (Jenuth *et al.*, 1996; Stewart *et al.*, 2008) and the occurrence of mitochondrial recombination in some animals (Piganeau *et al.*, 2004; Rubinoff *et al.*, 2006). Since mitochondrial genomes are predominantly maternally inherited, the resulting sequences may be sex-biased (Rubinoff *et al.*, 2006; Frézal and Leblois, 2008; Innocenti *et al.*, 2011). Selection pressures that differentially affect males and females are therefore not taken into account, such as differences in motility (Rubinoff and Sperling, 2004), feeding preference (Hulcr *et al.*, 2007), and the occurrence of genetic diseases (Frank and Hurst, 1996; Camus *et al.*, 2012). Additionally, the barcoding of insects infected with *Wolbachia* bacteria might lead to inaccuracies due to the effect that the symbionts have on reproductive functioning and the transferal of mitochondrial haplotypes in a population (Gerth *et al.*, 2011).

### **Differential rates of gene evolution**

Analogous to the uncertainties of the differential rate of mutation in the use of molecular clocks (Penny, 2005; Pulquério and Nichols, 2007; Weir and Schluter, 2008), the rate of evolution in the genes used for barcoding will vary between taxonomic groups (Rubinoff *et al.*, 2006). Mutation rates of genes in mitochondria are influenced by the unique asymmetrical mode of replication of the genome, where proximity to the site where replication begins is important (Gibson *et al.*, 2004). The lagging strand remains single-stranded until about two thirds of the leading strand has been replicated, which renders it more susceptible to mutations (Clayton, 2003; Rubinoff, 2006). Genome arrangements are not necessarily consistent across taxonomic groups owing to the vast array of genetic systems in existence, and to factors such as life history strategies (Rubinoff, 2006).

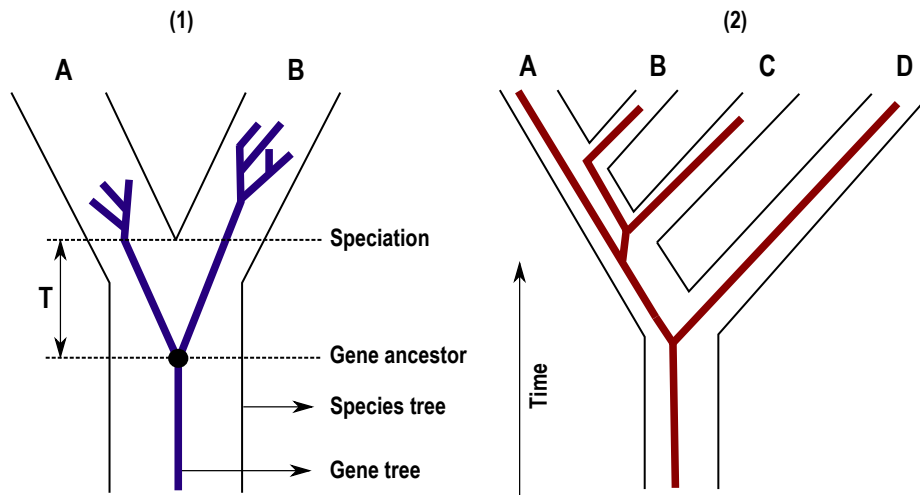
This could affect genetic barcoding through the size of the resulting barcode gap. A small or negative barcode gap results when there are too few mutational differences between groups to be able to tell them apart.

### **Introgression of mitochondrial genes**

Introgression, also referred to as interspecific gene flow, is the incorporation of genes from the gene pool of one species into that of another through hybridisation and backcrossing (Funk and Omland, 2003; Harrison and Larson, 2014). This leads to polyphyly and the incongruence between gene and species trees (Funk and Omland, 2003; Petit and Excoffier, 2009). The introgression of mitochondrial genes used in barcoding initiatives can result in misleading conclusions, as taxonomic boundaries can become blurred (Rubinoff, 2006; Petit and Excoffier, 2009; Ermakov *et al.*, 2015). Mitochondrial introgression is particularly problematic due to limited recombination, resulting in the persistence of introgressed gene regions in a genome (Funk and Omland, 2003). An important assumption in analysing intra- and interspecific variation is that the study species are monophyletic with regards to the gene region of interest (Funk and Omland, 2003). A number of studies have found that COI barcodes [alone] were unable to distinguish between species due to potential mitochondrial introgression and/or incomplete lineage sorting, such as in some cave-dwelling spiders (*Circurina* spp.) (Araneae: Dictynidae) (Paquin and Hedin, 2011), ground beetles (*Carabus* L.) spp. (Coleoptera: Carabidae) (Sota *et al.*, 2001; Raupach *et al.*, 2010), African fruit bats (*Pteropodidae* Gray) (Nesi *et al.*, 2011), swallowtail butterflies (*Papilio machaon* L.) (Lepidoptera: Papilionidae) (Sperling and Harrison, 1994), and hairstreak butterflies (*Calycopis* Scudder spp.) (Lepidoptera: Lycaenidae) (Cong *et al.*, 2017).

### **Incomplete lineage sorting**

As with introgression, incomplete lineage sorting (or ‘deep coalescence’) results in inconsistencies in topology between gene and species trees, such that the phylogeny of a particular gene does not match that of the species as a whole (Maddison, 1997; Nichols, 2001; Funk and Omland, 2003). Incomplete lineage sorting occurs when the coalescence of a gene back to a common ancestor pre-dates the speciation event (Fig. 1.3) (Maddison, 1997). As such, the ‘sorting’ of genes into the bifurcating lineages is not yet complete.



**Figure 1.3:** Gene trees within species trees, illustrating deep coalescence. (1) shows how the most recent common ancestor in the gene tree pre-dates that of the species tree by time  $T$ . (2) is another example of how the topology of a gene and species tree can differ. The species tree suggests that A and B share the most recent common ancestor, while the gene tree contrastingly shows that this is true for B and C. (Adapted from Maddison (1997); Nichols (2001); Edwards (2009)).

### Nuclear mitochondrial DNA segments

Nuclear mitochondrial DNAs ('numts'), also referred to as 'pseudogenes', are non-functional gene regions that have translocated from the mitochondrial to the nuclear genome (D'Errico *et al.*, 2004; Frézal and Leblois, 2008; Hazkani-Covo *et al.*, 2010; Leite, 2012). Pseudogenes were first discovered in mice (du Buy and Riley, 1967) and locusts (*Locusta migratoria* L.) (Orthoptera: Acrididae) (Gellissen *et al.*, 1983), and have since been found to occur frequently across a wide range of taxa (Lopez *et al.*, 1994; Zhang and Hewitt, 1996; Williams and Knowlton, 2001; Richly and Leister, 2004; Pamilo *et al.*, 2007); particularly in plants and arthropods (Gellissen *et al.*, 1983; Bensasson *et al.*, 2000, 2001; Buhay, 2009). One of the most extreme examples was found in the domestic cat, where nearly half of its mitochondrial genome was present in the nucleus (Lopez *et al.*, 1994)! It is believed that the transfer of mitochondrial genes to the nucleus is a very ancient process, and that pseudogenes might arise due to unsuccessful transfers between the two organelles (D'Errico *et al.*, 2004). Pseudogenes are problematic in DNA barcoding because these regions can be co-amplified in addition to the target mitochondrial gene (identifiable by differences in nucleotide composition, the presence of indels, point mutations, and in-frame stop codons) (Bensasson *et al.*, 2001; Song *et al.*, 2008; Moulton *et al.*, 2010; Ahmed and Jaffar Ali, 2015). Depending on the accumulation of mutations (where mutation rates and selection pressures differ between the mitochondrial and nuclear genomes), conservative primers

may preferentially amplify pseudogenes, resulting in paralogous sequences that have little comparative value and lead to the inference of incorrect phylogenetic relationships (Bensasson *et al.*, 2001; Song *et al.*, 2008; Moulton *et al.*, 2010; Leite, 2012). To further complicate matters, the number of copies and composition of pseudogenes can vary between individuals, where multiple independent transferral events could have occurred over time (Bensasson *et al.*, 2000, 2001). Song *et al.* (2008) found that the presence of pseudogenes in the COI barcodes for grasshoppers and crayfish lead to an overestimation of the number of species sampled by 2.8 and 3.6-fold, respectively. The authors concluded that the co-amplification of pseudogenes can be ‘disastrous’ for DNA barcoding if sequences are not screened properly, and if only one gene marker is relied upon. Despite the potential shortcomings of barcoding, the method has been successfully applied in numerous studies involving insects, representing groups such as the Heteroptera (Raupach *et al.*, 2014; Havemann *et al.*, 2018), Orthoptera (Hawlitschek *et al.*, 2017), Lepidoptera (Hausmann *et al.*, 2013; Razowski *et al.*, 2017; Spitsyn *et al.*, 2018), Coleoptera (Raupach *et al.*, 2010; Hendrich *et al.*, 2015), Diptera (Lin *et al.*, 2015), Ephemeroptera, Plecoptera, Trichoptera (Cordero *et al.*, 2017; Ćukušić *et al.*, 2017), and Hymenoptera (Schmidt *et al.*, 2017). If used appropriately, while remaining cognisant of its weaknesses, genetic barcoding can serve as a very valuable identification tool.

## 1.6 Inter-simple sequence repeats (ISSRs)

Inter-Simple Sequence Repeats (ISSRs) are sections of DNA located between microsatellites (=Simple Sequence Repeats (SSRs)) in an organism’s genome (Zietkiewicz *et al.*, 1994; Wolfe, 2005; Gaskin *et al.*, 2011) (Fig. B4). Microsatellites are sequences of repeated DNA motifs that usually occur randomly at multiple loci in a genome, range from one to six nucleotides in length, and display a high mutation rate (Gaskin *et al.*, 2011; Hoy, 2013). The ISSRs between these regions are particularly useful in detecting intraspecific variation due to their polymorphic nature (Abbot, 2001). In other words, each individual displays unique ISSR fragment sizes throughout their genome, which can be detected through the visualisation of banding patterns on an agarose gel and/or through capillary electrophoresis. ISSR primers typically consist of dinucleotide or trinucleotide base repeats (Wolfe, 2005). ISSR analysis was developed in 1994, and was originally used to differentiate between plant cultivars (Wolfe, 2005). Its application later extended to a wide range of other taxonomic groups, such as fungi (Kerrigan *et al.*, 2003), molluscs (Casu *et al.*, 2005, 2006), birds (Haig *et al.*,

2003; Wink, 2006), crocodiles (Machkour-M'Rabet *et al.*, 2009a), snakes (Guicking *et al.*, 2006; Nagy *et al.*, 2007), lizards (Joger *et al.*, 2007), tarantulas (Machkour-M'Rabet *et al.*, 2009b), sphingid moths (Lepidoptera: Sphingidae) (Hundsdoerfer and Wink, 2006), the diamondback moth (*Plutella xylostella* L.) (Roux *et al.*, 2007), lac insects (*Kerria* spp.) (Saha *et al.*, 2011), and silkworms (*Bombyx mori* L.) (Chatterjee and Mohandas, 2003). The technique has been used in the field of biological control in recent years, such as the genetic matching of invasive weeds to their native range (Paterson and Zachariades, 2013; Barker *et al.*, 2015; Sutton *et al.*, 2017) and the detection of cryptic species (Paterson *et al.*, 2016) and genetic diversity (Taylor *et al.*, 2011) within the insect agent *Eccritotarsus catarinensis* Carvalho (Hemiptera: Miridae), which is used for the biological control of water hyacinth. ISSR analysis is advantageous for the following reasons:

1. Primers are universal and can be used without prior knowledge of the target genomic sequence (Wolfe *et al.*, 1998; Gaskin *et al.*, 2011).
2. ISSRs are abundant in the genomes of a wide variety of taxonomic groups and are highly informative and reproducible.
3. The method is quick and cost-effective relative to other genotyping methods such as AFLP<sup>®</sup> (amplified fragment length polymorphism) and RFLP (restriction fragment length polymorphism) (Zietkiewicz *et al.*, 1994; Bornet and Branchard, 2001; Bornet *et al.*, 2004; AppliedBiosystems, 2014).
4. The mutation rate of microsatellite regions is rapid (Wan *et al.*, 2004; AppliedBiosystems, 2014).
5. Inherited microsatellite alleles are stable over multiple generations, and can be used at the population and individual level (AppliedBiosystems, 2014).

### 1.6.1 Caveats of inter-simple sequence repeats

Meudt and Clarke (2007) state that AFLPs tend to outperform ISSRs and RAPDs (random amplified polymorphic DNA) in terms of lower genotyping error and higher robustness and informativeness. A number of studies involving crop cultivars and other plants have compared different fragment analysis methods, and many have supported this statement (Jones *et al.*, 1997; McGregor *et al.*, 2000; Hodkinson *et al.*, 2002; Archak *et al.*, 2003; Garcia *et al.*, 2004; Behera *et al.*, 2008; Costa *et al.*, 2016).

Genotyping error refers to the differences between the genotypic data obtained independently from the same sample (Bonin *et al.*, 2004). The two overarching causes of such errors associated with 'band-based'

methods (including AFLPs and RAPDs) are allele homoplasy and scoring errors (Bonin *et al.*, 2004, 2007; Holland *et al.*, 2008).

### Homoplasy

The term ‘homoplasy’ means that the same characteristic has evolved independently in two or more otherwise different entities/species. Allele homoplasy thus refers to instances where fragments appear at the same position on a gel or an electropherogram, but are non-homologous (Bonin *et al.*, 2007; Simmons *et al.*, 2007). This phenomenon appears to increase in frequency in smaller fragments and in dense profiles, and when taxonomic distance increases; such that its effect on intraspecific comparisons is negligible (Bonin *et al.*, 2007). Conversely, a particular locus may have undergone a mutation, insertion or deletion, resulting in the appearance of two or more independent loci that are actually linked (Simmons *et al.*, 2007). Simmons *et al.* (2007) also mention the concerns around the origins of the amplified fragments (i.e. whether they are from the nuclear or mitochondrial genome, or whether they are in fact contaminants from hosts, parasites, or symbionts).

### Error rates

Data scoring is considered to be the main source of errors in fragment analysis procedures due to the high level of subjectivity involved at various stages of the analysis process (Bonin *et al.*, 2004). Scoring errors for ISSRs can be largely overcome through the replication of each sample at the polymerase chain reaction (PCR) step. The most important determining factors in the comparison of replicate samples include:

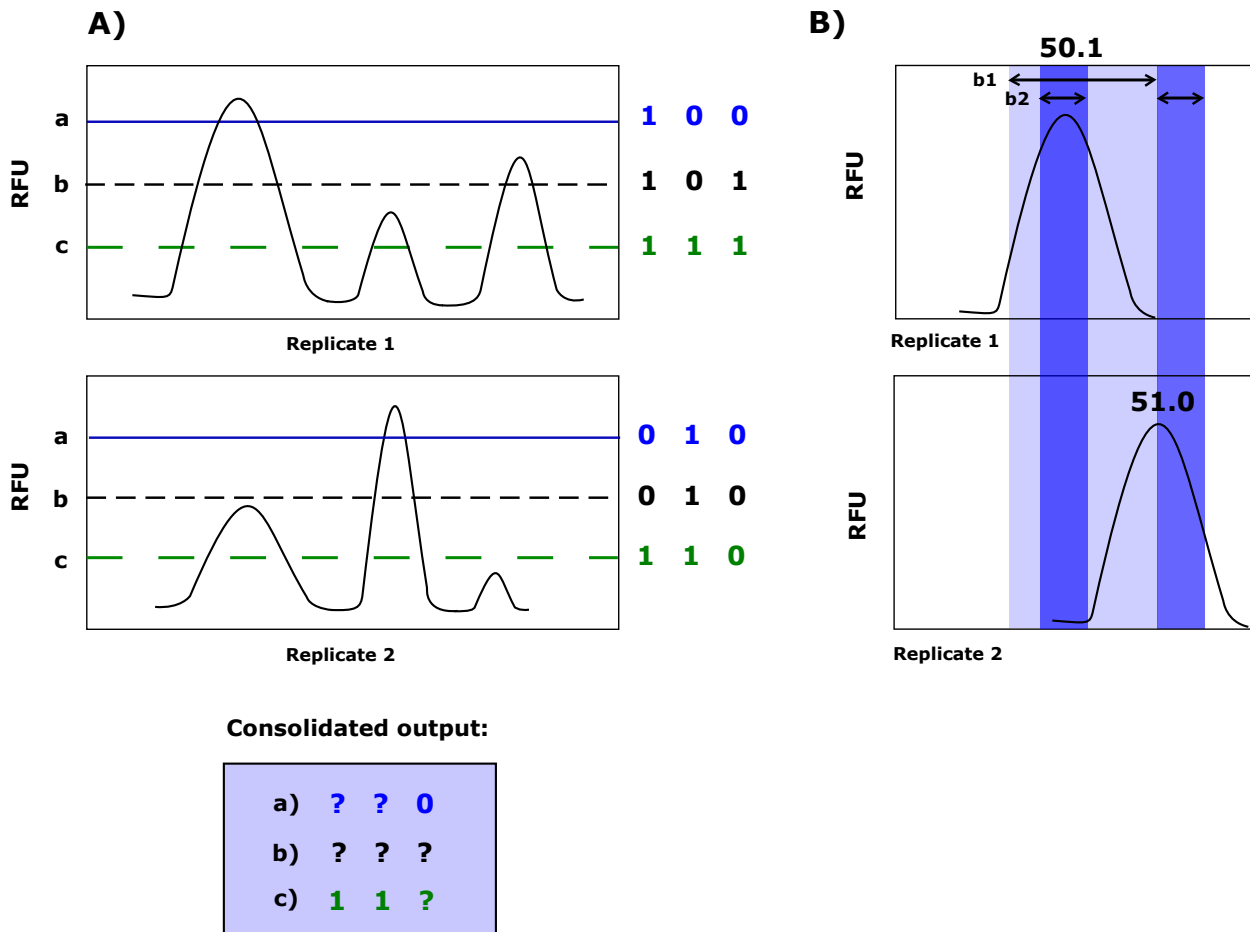
1. Differences in band intensity and what the peak height threshold is set at (i.e. what level of fluorescence is considered background noise, and how the number of false positive and negative calls can be minimised?).
2. The bin width applied to score the presence and absence of bands (i.e. how strict or lenient should one be when determining whether a band occurs at the same locus in a replicate pair? What level of leeway should there be? If the bin width is too large, otherwise separate characters will be grouped as one, whereas a bin width that is too small will result in the scoring of numerous characters which are actually only representative of one locus) (Fig. 1.4).

Error rates are typically calculated as the average ratio between the number of mismatches and the total number of instances where a band was present, and absent, in both replicates (referred to as the Euclidean

error rate) (Pompanon *et al.*, 2005; Bonin *et al.*, 2007). It is suggested that an alternative error rate (the Jaccard error) should be calculated to complement the Euclidean error value. The Jaccard error calculation does not consider the shared absence of a band, as this absence may not be biologically meaningful (Holland *et al.*, 2008). AFLP Euclidean error rates typically range between 2 and 5%, but Bonin *et al.* (2007) suggest that it should ideally not exceed 1%. Holland *et al.* (2008), however, suggest that inter-study comparisons of error rates are not always reliable, as they encompass both PCR and scoring errors, and varying levels of divergence and data set size. Their study, for example, yielded AFLP error rates of 9 to 18% and 6 to 13% for sweet potato (*Ipomoea batatas* (Convolvulaceae)) and *Ourisia* (Plantaginaceae), respectively. Additionally, most of the studies reporting genotyping errors refer to AFLP techniques, and so it is unclear to what degree ISSR error rates are comparable. Surveys conducted by Guichoux *et al.* (2011) and Bonin *et al.* (2004) found that only between 6 and 26% of papers, respectively, dealing with microsatellites reported error rates.

### 1.6.2 ISSRs as an identification tool

Due to the significantly higher mutation rate in microsatellites compared to other DNA regions (Zietkiewicz *et al.*, 1994; Reddy *et al.*, 2002), ISSRs can offer higher resolution results that reveal greater differences between groups. This could be valuable in cases where traditional DNA barcoding regions are not sufficient in differentiating between closely related taxa, such as intraspecific lineages and different population groups. Fragment analysis methods, such as ISSRs, can therefore be used as molecular barcodes through the creation of unique fingerprint profiles (see for example Galbács *et al.* (2009), Adhikari *et al.* (2015), Li *et al.* (2018), and Zhao *et al.* (2019)).



**Figure 1.4:** Potential sources of scoring errors in band-based genotyping methods: A) Determining the optimal threshold value to apply when scoring the presence or absence of a peak. The illustration shows three RFU (Relative Fluorescence Units) threshold levels (a, b and c), the resulting binary output at each, and the consolidated binary output representing a replicate pair (where a 1 and 0 or 0 and 1 outputs a ‘?’; the shared presence of a band outputs a ‘1’ and the shared absence a ‘0’). B) The effect of bin width on the scoring of peaks. Replicate 1 shows a band of 50.1 base pairs (bp) in size, and replicate 2 shows one of 0.9 bp larger. With a bin width set at b1, both will be scored as present at the same locus, but at a width of b2, these bands will be scored as two separate loci.

## 1.7 Aims

The aims of this Master’s thesis were to:

1. Obtain DNA and molecular barcode data to create phylogenies for the Dactylopiidae.
2. Test the accuracy of these barcode databases.
3. Develop protocols to allow for the accurate and quick identification of species, and more importantly, intraspecific lineages within *Dactylopius opuntiae* and *D. tomentosus*.

To date, a number of distinct genetic lineages of *D. tomentosus* have been determined (e.g. Mathenge

*et al.* (2009a, 2010b)), but not for *D. opuntiae* (namely the ‘ficus’ and ‘stricta’ lineages). Filling this knowledge gap is very valuable to the biological control of invasive Cactaceae around the world, and offers a number of important practical applications.

# Chapter 2

## Materials and Methods

This document contains two appendices: A and B. Figures and Tables in Appendix A begin with the prefix ‘A’, and Section numbers with ‘5’. The Figures and Tables in Appendix B begin with the prefix ‘B’, and Section numbers with ‘6’.

### 2.1 Sourcing of samples

Tables A1 and A2 contain detailed lists of all the samples used in this study. Outgroup and additional ingroup sequences were obtained from Genbank to supplement the dataset, and are listed in Table A3.

#### 2.1.1 South Africa

##### Wild populations

*Dactylopius opuntiae* samples were collected from wild populations on *Opuntia ficus-indica* and *Opuntia engelmannii* from various locations in the Eastern Cape Province. These are known to be the ‘ficus’ lineage.

##### Uitenhage Mass Rearing Facility

The Uitenhage Mass Rearing Facility in the Eastern Cape Province rears six insect agents for the control of invasive Cactaceae in the country. Four of these are *Dactylopius* species, and were included in the analyses of this project. These were *D. austrinus*, *D. tomentosus* ‘imbricata’, *D. ceylonicus*, and *D. opuntiae* ‘stricta’ (more information is available at the Centre for Biological Control (CBC) Website, at <https://www.ru.ac.za/centreforbiologicalcontrol/massrearing/uitenhagemassrearingfacility/>).

As there are wild populations of the ‘ficus’ *D. opuntiae* lineage on *O. ficus-indica* cacti in the surrounding vicinity, it is a concern as to whether this lineage is hybridising with the desired ‘stricta’ lineage reared inside

the facility. It is known that these two lineages can hybridise, which affects the host-specificity of offspring. ‘Stricta’ samples were collected from the facility and compared to original ‘stricta’ source population samples and to ‘ficus’ specimens to determine whether hybridisation had occurred. Individuals of the ‘stricta’ lineage from the original stock imported from Australia were kept by Hildegard Klein and John Hoffmann (biological control specialists), and were used as a benchmark for the identification of this lineage.

### ***Dactylopius tomentosus* ‘cholla’**

The *Dactylopius tomentosus* ‘cholla’ lineage was sourced from *Cylindropuntia fulgida* var. *mammilata* in the Kirstenbosch Cape Town Botanical Gardens. Additional samples collected from *C. fulgida* var. *mammilata* in Jansenville in the Eastern Cape Province required identification confirmation, and were included in the analyses.

### **Kruger National Park**

*Opuntia stricta* is the most widespread IAP in the Kruger National Park (KNP), covering approximately 35 000 ha surrounding Skukuza (Lotter and Hoffmann, 1998; Foxcroft *et al.*, 2004). It was first recorded in the Skukuza staff village in 1953, and spread rapidly within the subsequent decades (Lotter and Hoffmann, 1998). *Dactylopius opuntiae* was released on three separate occasions between 1990 and 1995, but did not establish on the weed (Lotter and Hoffmann, 1998; Foxcroft and Hoffmann, 2000). This was because it was the ‘ficus’ lineage, which does not establish on *O. stricta* (Volchansky *et al.*, 1999). The ‘stricta’ lineage was subsequently imported from an Australian stock and released in the KNP in 1997, and has provided a substantial level of control (Lotter and Hoffmann, 1998; Foxcroft and Hoffmann, 2000). *Dactylopius opuntiae* samples were obtained from Skukuza in the KNP in 2018 in order to confirm whether they were still of the original ‘stricta’ lineage released in 1997.

### **2.1.2 Namibia**

*Dactylopius opuntiae* samples were collected from *O. ficus-indica* and *O. stricta*. These were expected to be the ‘ficus’ lineage, as this was the only lineage intentionally released in the country in 1975 and 1980 (Brown *et al.*, 1985; Paterson *et al.*, 2019). The first release in 1975 was for the control of *O. stricta*, but the ‘ficus’ lineage did not establish (Brown *et al.*, 1985) due to it being the incorrect agent. This

is another example of the importance of correct identification. Additionally, two samples collected on an unidentified *Harrisia* sp. in Windhoek were included in the present analysis. It was unexpected to find *D. opuntiae* ‘ficus’ on a *Harrisia* cactus, and so further identification was required.

### 2.1.3 The USA

*Dactylopius* populations were collected from ten different *O. engelmannii* lineages across the southern states of Arizona, New Mexico, Texas, and California in 2017 (Byrne *et al.*, 2018), and were identified morphologically as *D. opuntiae*. These are currently in quarantine at the University of the Witwatersrand in Johannesburg, and are being tested for their use as additional biological control agents for *O. engelmannii*. Samples from each lineage were included in the present genetic analysis. Additional confirmed *D. confusus* specimens were sourced from the Entomology Department at the University of Arizona, and included in the analysis for comparison. This is because *D. opuntiae* and *D. confusus* are frequently confused, as they are often found on the same individual host plant. Genetic analyses could therefore confirm the morphological identifications made for these ten lineages.

### 2.1.4 Australia

There are eight naturalised *Cylindropuntia* species in Australia; namely *C. fulgida* var. *mamillata* ((D. C.) Backeberg), *C. imbricata* (Haworth) F. M. Knuth, *C. kleimiae* (D. C.) F. M. Knuth, *C. leptocaulis* (D. C.) F. M. Knuth, *C. prolifera* (Engelm.) F. M. Knuth, *C. rosea* (D. C.) Backeberg, *C. spinosior* (Engelm.) F. M. Knuth, and *C. tunicata* (Lehm.) F. M. Knuth (Jones *et al.*, 2015). All of these species negatively impact the environment, agricultural practices, and recreational activities (Lloyd and Reeves, 2014; Jones *et al.*, 2015). The Australians have imported 20 *D. tomentosus* lineages into the country to serve as potential biocontrol agents (Jones *et al.*, 2018), of which the ‘bigelovii’, ‘acanthocarapa x echinocarpa’, ‘cylindropuntia sp.’, ‘imbricata’, ‘cholla’, and ‘californica’ lineages were obtained from Michael Day at the Department of Agriculture, Fisheries and Forestry, (Biosecurity Queensland) for inclusion in this project. These lineages are the most damaging to target weeds, and are in the process of, or have already been, released. Additional *D. austrinus*, *D. ceylonicus*, and *D. opuntiae* specimens were included as well, which are commonly used biological control agents in Australia for Cactaceae, and have been established in the country for many years.

### 2.1.5 Saudi Arabia

*Dactylopius opuntiae* samples collected from *Opuntia stricta* were sourced from four localities in the Jazan Region of Saudi Arabia; namely Aldyer, Jazan City, Hurob, and Alaredah. Confirmation was required that these belonged to the ‘stricta’ lineage, because this particular lineage provides the greatest level of control for *O. stricta* weeds.

## 2.2 Sample preservation and DNA extraction

All specimens were preserved in 95% ethanol and stored at -18°C. Microscope slides of selected specimens were prepared according to the methods of Ben-Dov and Hodgson (1997). Buffer ATL (tissue lysis buffer) (Qiagen<sup>©</sup>) and Proteinase K (Qiagen<sup>©</sup>) were used to digest insect tissue, followed by a standard salt extraction procedure (Bruford *et al.*, 1992) (See Section 5.1 for the detailed protocol). DNA pellets were resuspended in 100  $\mu$ L AE (elution) buffer (Qiagen<sup>©</sup>) and stored at -18°C. The concentration and purity of DNA extracts were ascertained through the use of a ThermoScientific Nanodrop 2000 spectrophotometer. When present, carminic acid residues in the extracts affected spectrometry and DNA concentration readings, but did not inhibit the successful amplification of target fragments.

## 2.3 DNA sequencing

### 2.3.1 Gene regions

Three gene regions were amplified; namely the mitochondrial cytochrome oxidase subunit 1 (COI) region and 12S ribosomal (12S rRNA) region, and the nuclear 18S ribosomal (18S rRNA) region. Mitochondrial DNA is maternally inherited, and is highly conserved across the insect orders and bilaterian animals as a whole (Hoy, 2013; Cameron, 2014). This high level of conservation makes it particularly useful for phylogenetic studies (Lunt *et al.*, 1996; Hoy, 2013; De Mandal *et al.*, 2014). Another useful property of mtDNA is that different gene regions mutate at different rates, allowing for greater analytical potential (Hoy, 2013). The third position nucleotides of the COI gene display a high rate of base substitutions (‘wobble’), leading to a mutation rate about three times faster than that of the 12S and 16S rDNA genes (Knowlton and Weigt, 1998). This

rapid rate of gene evolution has been found to be useful in using genetic barcoding to discriminate between phylogeographic lineages within the same species (Cox and Hebert, 2001; Gentekaki and Lynn, 2012). The 18S gene is one of the slowest-evolving nuclear sequences, and is useful for assessing the ancestral relationships between organisms (Hillis and Dixon, 1991). Campana *et al.* (2015) did not find any variation between the 18S markers across *D. coccus* lineages, and hence discarded them from their study. The present study however used this gene because it involved a multi-species analysis.

### 2.3.2 Polymerase Chain Reaction (PCR)

Forward and reverse primers were used according to the methods of Park *et al.* (2010), Campana *et al.* (2015), and Mathenge *et al.* (2015) (Table 2.1).

**Table 2.1:** Primers used for the amplification of the COI, 12S, and 18S gene regions in *Dactylopius* species.

Gene	Primer pair (5' → 3')	Reference
COI-A	<b>Forward (PcoF1)</b> CCTTCAACTAATCATAAAAATATYAG	Park <i>et al.</i> (2010)
	<b>Reverse (LepR1)</b> TAAACTTCTGGATGTCCAAAAAATCA	
COI-B	<b>Forward (DTOMf)</b> TCCGRATAGAACTWATAAAYACYAA	Mathenge <i>et al.</i> (2015)
	<b>Reverse (HCO2198)</b> TAAACTTCAGGGTGACCAAAAAAATCA	
12S	<b>Forward (12S-F)</b> AAGAGTGACGGGCRATTTGTACATA	Campana <i>et al.</i> (2015)
	<b>Reverse (12S-R)</b> GTGCCAGCAGTWGCGGTTA	
18S	<b>Forward (18S-F)</b> CTGGTTGATCCTGCCAGTAG	Campana <i>et al.</i> (2015)
	<b>Reverse (18S-R)</b> CCGCGGCTGCTGGCACCAGA	

The PcoF1 forward primer designed by Park *et al.* (2010) binds upstream of the COI gene, in the tRNA domain, as this is a more conserved region. This increases the chance of amplification in groups that do not amplify with universal primers, as is generally found in the scale insects. This was true for the Dactylopiidae, as the universal HCO2198 and LCO1490 COI primers were tested here, and did not work for any species. The two COI primer pairs used here (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*, *D. opuntiae*, and *D. confusus*; while COI-B only worked for *D. tomentosus*, as it was particularly designed by Mathenge *et al.* (2015) for this species. Samples were prepared according to the protocol in Table 2.2, and amplified in a Bio-Rad T100 Thermal Cycler<sup>TM</sup> set according to the protocol given in Table 2.3.

**Table 2.2:** PCR protocol for the 12S, 18S, COI-A, and COI-B genes in *Dactylopius*.

Constituent	Quantity ( $\mu\text{L}$ )		Concentration
iTaq <sup>TM</sup> Master Mix	12.5		
	<b>Gene</b>		
	<b>12S</b>	<b>18S</b>	<b>COI-A &amp; COI-B</b>
Forward primer	3	3	2
Reverse primer	3	3	2
ddH <sub>2</sub> O	4.5	4.5	6.5
DNA sample	2	2	2
<b>Total</b>	<b>25</b>		$\sim 50\text{-}150 \text{ ng}/\mu\text{L}$

**Table 2.3:** PCR protocol for the 12S, 18S, COI-A, and COI-B genes.

Cycles	Step	Temperature ( $^{\circ}\text{C}$ )	Time
1	Initial denaturation	94	5 min
	Denaturation	94	40 s
45	Annealing	52 (12S); 58 (18S); 43 (COI-A); 50 (COI-B)	35 s
	Extension	72	45 s
1	Final extension	72	10 min
1	Hold	12	$\infty$

iTaq<sup>TM</sup> Universal SYBR<sup>®</sup> Green Supermix (Bio-Rad Laboratories Inc., USA) provided better amplification than TopTaq<sup>®</sup> (Qiagen, Germany). The 12S, 18S, and COI genes yielded products approximately 450-500, 630-650, and 600-630 base pairs long. The 12S primers had a tendency to produce double bands for some samples (the target fragment plus a fragment approximately 1500 bp in size. See Fig A1). Touchdown PCR (TD-PCR) was applied in such cases according to the protocol in Table 2.4.

**Table 2.4:** 12S Touchdown-PCR protocol that addressed double-band amplification.

Cycles	Step	Temperature (°C)	Time
1	Initial denaturation	94	5 min
	Denaturation	94	30 s
5	Annealing	54	30 s
	Extension	72	30 s
	Denaturation	94	30 s
35	Annealing	51	40 s
	Extension	72	40 s
1	Final extension	72	8 min
1	Hold	12	∞

### 2.3.3 Gel electrophoresis

Gels were prepared using 100 mL TBE buffer (1X), 1 g Seakem<sup>TM</sup> agarose (1%) and 10  $\mu$ L ethidium bromide. A 2  $\mu$ L aliquot of loading buffer (PCRBIO Loading Buffer A) was added to 5  $\mu$ L of each DNA sample. A 2  $\mu$ L DNA ladder (PCRBIO Ladder IV, PCRBIO SYSTEMS) served as a standard. Gels were run at 90 V for 40 minutes using a Bio-Rad Powerpac<sup>TM</sup>. Bands were visualized on a Bio-Rad Geldoc<sup>TM</sup> molecular imager system and photographed using Image Lab<sup>TM</sup> software. Amplified PCR products were sent to Macrogen Inc. in the Netherlands for purification, and for sequencing in the forward direction in all cases.

### 2.3.4 Sequence alignment

Chromatograms were opened in Chromas v2.6.4 (Technelysium Pty Ltd.), where the beginning and ends of sequences were trimmed accordingly ( $\pm$  50 - 70 bp). These were subsequently opened in BioEdit v7.0.5 (Hall, 1999), and their base calls verified against their corresponding chromatograms. Sequences were aligned online using MAFFT v7 (Kato *et al.*, 2017). The MAFFT alignment method was chosen based on the review by Morrison (2006), which found this method to be the most accurate and consistent compared to the ProAlign, ClustalW, T-coffee, Muscle, and Clustal methods.

As an additional analysis, the 12S rRNA sequences were aligned according to their predicted secondary structures using R-Coffee (Moretti *et al.*, 2008; Wilm *et al.*, 2008) on the [T-Coffee server](#) (Notredame *et al.*, 2000; Di Tommaso *et al.*, 2011). Sequences that did not show high consistencies were removed from the alignment. Final sequences were submitted to Genbank via the online submission portal available at <https://submit.ncbi.nlm.nih.gov/>. See Section 5.2.1 for more details regarding the Genbank submission process.

### 2.3.5 Construction of phylogenetic trees

jModelTest v2.1.10 (Guindon and Gascuel, 2003; Darriba *et al.*, 2012) was used to select appropriate evolutionary models for the 12S, 18S, and COI gene regions, followed by the use of Bayesian Inference and Maximum Likelihood methods to create phylogenies for each. A concatenated phylogeny was made for the 12S and 18S regions. 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*.

Maximum Likelihood (Felsenstein, 1981) phylogenies were created using GARLI v2.01 (Zwickl, 2006), and Bayesian phylogenies were created using MrBayes v3.2.6 (Huelsenbeck and Ronquist, 2001) via the CIPRES Science Gateway portal (Miller *et al.*, 2010) (<http://www.phylo.org/>). Trees were viewed in FigTree v1.4.3 and edited in Inkscape v0.92.3.

#### Bayesian Inference and Maximum Likelihood

MrBayes was run using random starting trees, four chains (three hot and one cold) and set to 50 million generations. Trees were sampled every 1000 generations with a burnin of 1250 (see Section 5.9 for the MrBayes Nexus files used for each gene region). Tracer v1.7 was used to check for MCMC convergence (Rambaut *et al.*, 2018).

Protein-coding COI sequences were assessed for substitution saturation at positions 1, 2, 3, and 1 & 2 using the tests created by Xia *et al.* (2003) and Xia and Lemey (2009) in DAMBE v6 (Xia, 2018). It is important to check for codon saturation, because the actual genetic distances in a dataset could be significantly underestimated due to multiple substitutions having taken place (Philippe *et al.*, 2011). Sequences were ensured to be in the longest open-reading frame using BioEdit v7.0.5.3. GARLI and CONSENSE were used for the construction of Maximum Likelihood phylogenies via CIPRES. One thousand bootstrap repeats were run with two search repetitions.

Topologies of the individual gene trees for the 12S and 18S gene regions were compared using the online congruency index test ‘Icog’, developed by de Vienne *et al.* (2007). To do this, a neighbour-joining tree was created for each gene region in MEGA v7 (Kumar *et al.*, 2016) using the default settings. A significant p-value in ‘Icog’ supports congruency, and thus the concatenation of the two relevant genetic datasets.

## Haplotype networks in POPART

POPART v1.7 (Leigh and Bryant, 2015) was used to create 12S and COI haplotype networks for *D. opuntiae* and *D. confusus* using the TCS network method (Clement *et al.*, 2000). *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 was therefore not included. See Section 5.3.2 for the relevant file input formatting for POPART.

### 2.3.6 Distance-based testing of barcodes

The accuracy of barcode sequences were tested using the SPIDER package (Brown *et al.*, 2012) in R v3.6.1 (R Core Team, 2013) at both the species and lineage level. This entailed the use of the Nearest Neighbour (NN), Threshold Identification (TID), and Best Close Match (BCM) methods. The threshold percentage used in TID was optimised to minimise cumulative error (false negatives + false positives). This was done by testing a range of threshold values from 0.0001% to 2.5% in increments of 0.005% using the threshold optimisation function in the package. Pairwise distances were calculated using the K80 substitution model, using the `dna.dist` function in the R ‘Ape’ package (Paradis *et al.*, 2004). The genetic barcoding gap was calculated and plotted using the SPIDER package by calculating the largest intraspecific and the smallest interspecific genetic distance for each sequence (Meier *et al.*, 2006). See Section 5.7.1 for the R code used to implement these tests.

#### Nearest Neighbour (NN)

The NN algorithm was used to compare the genetic distance of each sequence of interest (treating it as a query sequence) to every other sequence in the relevant dataset. This method found the sequence that was most similar to it, and then checked whether this most similar sequence was conspecific or not. This was based upon the predefined species names, which returned a ‘true’ or ‘false’ result. In this way, the accuracy of each species’ identification was tested and converted into a percentage of positive and negative outcomes.

#### Threshold ID (TID)

The TID algorithm was used to find all the sequences that were within a genetic distance threshold of the query sequence, and returned ‘correct’, ‘incorrect’, or ‘no id’ based on whether those matches were

conspecific to the query sequence or not. The distance threshold was 1% by default, but was set to an optimal value that minimised false positive and false negative results.

### **Best Close Match (BCM)**

The BCM algorithm was used in addition to the NN and TID, but instead of narrowing down all the matches within the genetic distance threshold, it only compared to the single nearest neighbour. It also returned ‘correct’, ‘incorrect’, or ‘no id’ as output. The ambiguous output ‘no id’ was returned when more than one sequence matched the query, where these matches included both conspecific and allospecific individuals.

### **Scoring of the barcode algorithms**

As per the methods of Birch *et al.* (2017) and the R SPIDER documentation (Brown *et al.*, 2012), species and lineage identifications were considered:

1. “True” in NN when the closest individual to the query was conspecific, and “correct” in BCM and TID analyses when all individuals with the closest distance to the query were conspecific and fell within the threshold applied.
2. “Ambiguous” in BCM analyses when different allospecific individuals shared the closest distance to the query and were within the threshold value or in TID analyses when different allospecific individuals were within the threshold value.
3. “No identification” in BCM and TID analyses when individuals were genetically more distant to the query than the threshold value.
4. “False” in NN when the closest individual to the query was allospecific, and “incorrect” in BCM analyses when allospecific individuals shared the closest distance to the query and were within the threshold value or in TID analyses when all individuals within the threshold value were allospecific.

## 2.4 Inter-simple Sequence Repeats (ISSRs)

### 2.4.1 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'-FAM<sup>TM</sup> fluorescent dye. PCRs were run in 20  $\mu$ L reactions, consisting of 10  $\mu$ L iTaq<sup>TM</sup>, 1.5  $\mu$ M primer and 1  $\mu$ L DNA template (50 - 150 ng/ $\mu$ L). The PCR protocol was adapted from that of Saha *et al.* (2011) and Silva *et al.* (2013), and carried out as per Table 2.5. A replication of each sample was conducted in a different PCR machine in order to validate the occurrence of bands/peaks (Taylor *et al.*, 2011; Sutton *et al.*, 2017). As a preliminary validation step, 5  $\mu$ L PCR product was run on a 1.5% agarose gel at 6 V/cm for 3 hours (Fig. A2). Samples were subsequently sent to the Central Analytical Facilities (CAF) division in Stellenbosch, South Africa, where DNA fragment analysis was performed using capillary electrophoresis (Applied Biosystems Inc., 3130 genetic analyser, GS500LIZ size standard).

**Table 2.5:** PCR protocol for the ISSR 809 and ISSR 826 primers for *Dactylopius* spp.

Cycles	Step	Temperature ( $^{\circ}$ C)	Time
1	Initial denaturation	95	7 min
	Denaturation	95	45 sec
38	Annealing	44 (ISSR 809); 48 (ISSR 826)	45 sec
	Extension	72	2 min
1	Final extension	72	7 min
1	Hold	12	$\infty$

### 2.4.2 ISSR data processing and analysis

#### Electropherograms to binary data

Electropherograms were analysed in Genemarker<sup>®</sup> v2.7.4 (SoftGenetics<sup>®</sup>) (Fig. A6). See Section 5.3.1 for more details regarding the settings applied. Binary data from GeneMarker<sup>®</sup> were opened in RawGeno v2.0 (Arrigo *et al.*, 2012). The following parameters were found to be the most efficient, while remaining conservative:

1. Maximum and minimum bin width of 1 and 0.5 base pairs (bp), respectively.
2. Scoring from 100 to 500 bp.
3. Low relative fluorescent units (RFU) of 100, as recommended by AppliedBiosystems (2014) for data produced by 3130 Series instruments.

The resulting binary matrix was organised in Microsoft Excel<sup>®</sup> and converted to a pairwise distance matrix in R using Jaccard's distance transformation. This transformation was applied because it does not consider the shared absence of bands as being biologically meaningful (Fig. A4) (Jaccard, 1908). See Section 5.7.2 for the relevant R code used to calculate distance matrices applying Jaccard's transformation. The Jaccard's distance ( $dJ_i$ ), and similarity ( $J_i$ ) index calculations are shown below, where  $f_{11}$  represents the total number of times that a band occurred at the same place in both samples,  $f_{00}$  represents the shared absence of bands, and  $f_{10}$  and  $f_{01}$  represents the number of times that a band was present in only one of the two sample replicates. Jaccard's distance is equal to 1 - Jaccard's similarity ( $J_i$ ).

$$dJ_i = \frac{f_{01} + f_{10}}{f_{01} + f_{10} + f_{11}} = 1 - J_i$$

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

## Grouping methods

The transformed distance matrix was subsequently used to create a hierarchical clustering dendrogram using the unweighted pair group method with arithmetic mean (UPGMA, 1000 bootstraps) using the 'pvclust' R package (Suzuki and Shimodaira, 2006). See Section 5.7.3 for the R code used to construct this dendrogram. Further ISSR clustering and statistical analyses were applied to the *Dactylopius opuntiae* 'ficus' and 'stricta' lineages in particular, as the traditional DNA barcoding regions could not differentiate between them. Kruskal's non-metric multidimensional scaling (nMDS) method was applied using the isoMDS function in the MASS package in R (Venables and Ripley, 2002). See Section 5.7.4 for the relevant R code used to create the nMDS plot. Dimensions  $K = 2$  and  $K = 3$  were tested, and the one yielding the lowest stress value was selected. The use of an nMDS ordination method was confirmed using stress and Shepard plots. An ANOSIM

(ANalysis Of SIMilarity) was applied to lineages and populations using the ‘vegan’ package in R (Oksanen *et al.*, 2018), using 999 permutations (Clarke, 1993). Pairwise p and R-values were obtained from PAST v3 (Hammer *et al.*, 2001). In addition to the ANOSIM, a Permutational Multivariate Analysis Of Variance (PERMANOVA) using distance matrices was used, applying the `adonis` function in the ‘vegan’ package, with 999 permutations. This test is less sensitive to within-group heterogeneity (Anderson, 2014). Post-hoc pairwise comparisons were obtained via the `pairwise.perm.manova` function in the ‘RVAideMemoire’ package (Hervé, 2015). See Section 5.7.6 for the R code used to output the ANOSIM and PERMANOVA test results. Data were also imported to SplitsTree4 v4.15.1 (Huson and Bryant, 2006), and used to create NeighborNet trees based on distance matrices that were transformed using Jaccard’s distance index. NeighborNet trees were chosen because Huson and Bryant (2006) suggest that this method produces highly resolved networks. See Section 5.3.3 for the SplitsTree file input format.

## STRUCTURE

STRUCTURE v2.3.4 (Pritchard *et al.*, 2000; Evanno *et al.*, 2005) was used to analyse the genetic clustering of *D. opuntiae* samples. The program uses Bayesian analysis and the Markov Chain Monte Carlo (MCMC) estimation method to assign individuals to genotypic cluster groups ( $K$ ) (Pritchard *et al.*, 2000). STRUCTURE was run twice; once where the population origins for samples were given as priors (‘PopData’), and once where they were not. See Section 5.3.4 for details regarding the specific settings used. The results obtained from STRUCTURE were uploaded to STRUCTURE SELECTOR (Li and Liu, 2018), which finds the best estimate of the number of genetic cluster groups ( $K$ -value) present. The Puechmaille method was preferred for this analysis, as it was found to outperform previous methods for datasets containing both even and uneven sample sizes (Puechmaille, 2016). Puechmaille (2016) developed four statistical tests to estimate the optimal  $K$ -value for a dataset, namely ‘MedMeaK’, ‘MaxMeaK’, ‘MedMedK’, and ‘MaxMedK’. The present analysis set the threshold to range between 0.5 and 0.8, as per the recommendations of Puechmaille (2016). The  $K$ -estimation according to the conventional Evanno method (Evanno *et al.*, 2005) was also reported. CLUMPAK software (Kopelman *et al.*, 2015) was used to visualise results.

## 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).

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

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

Due to the JE not taking the shared absence of bands into account, it will be inflated if the data contain numerous band-absent loci.

## 2.5 Graphical User Interface (GUI) Applications

### 2.5.1 R Shiny Applications

R Shiny applications (RStudio, Inc, 2013) are hosted on Amazon’s Web Services online platform, and the front end can be accessed by any user via a unique url link. Running the program only requires an Internet connection, and does not rely on any pre-installed R software or related packages. Each Shiny application comprises of a user interface object (ui.R) and a server function (server.R), which are uploaded to a profile on <https://www.shinyapps.io/>. Three such applications were created and are described below. A help file accompanies each application.

### 2.5.2 Barcode Testing

An application that allows a user to apply barcode testing algorithms to a set of aligned FASTA sequences (Fig. A13) is available at <https://clarkevansteenderen.shinyapps.io/BarcodeTester/>. The barcode tests were taken from the SPIDER package developed by Brown *et al.* (2012). The user can also view and download a heat map representation of a distance matrix, plots for base frequencies and GC content (using the ‘Ape’ package (Paradis *et al.*, 2004)), and a UPGMA clustering tree in the case of binary data.

### 2.5.3 Binary data processing: ‘BINMAT’

A second GUI application (Fig. A14) was created in order to rapidly consolidate replicate pairs in a binary matrix containing fragment analysis data from a dominant genetic marker such as an ISSR or AFLP analysis (i.e. it requires already-scored data from a program such as GeneMarker<sup>®</sup>). The algorithm combines replicates such that if a band was only present in one replicate (“0” and a “1”), it is treated as ambiguous (scored as a “?”). The dual presence or absence of a band remains a “1” or “0”, respectively. The application also calculates the average Euclidean and Jaccard error rates with standard deviations, and outputs the average, minimum, and maximum number of peaks and their standard deviations. An option is provided for constructing a UPGMA clustering tree to preview the clustering pattern of the data. An additional feature allows the user to upload their consolidated matrix to create an interactive nMDS plot that can be downloaded as a .svg file. Options are available to view scree and Shepard plots for the data, download the distance matrix, and remove samples with total peak numbers below a specified threshold value. These filtered datasets can be downloaded and re-uploaded to create new nMDS plots. The BINMAT application is accessible at <https://clarkevansteenderen.shinyapps.io/BINMAT/>. It is also available via GitHub, where users can access the GUI by typing `> shiny::runGitHub("BinMat", "CJMvS")` into the R console on their local machine.

### 2.5.4 Identification of query sequences: ‘Dacty-ID’

The third GUI application allows the user to paste a query nucleotide sequence into a search box, which is then aligned with sequences in a pre-selected database (Fig. A15). The databases are the sequences for the 12S, 18S, and COI gene regions (COI-A and COI-B) obtained in this project. Alternatively, the user can upload their own set of sequences to serve as a database. A phylogenetic tree is subsequently created, where the position of the query sequence is highlighted in red relative to the database. The Dacty-ID application is available at [https://clarkevansteenderen.shinyapps.io/Dactylopius\\_ID\\_version\\_1/](https://clarkevansteenderen.shinyapps.io/Dactylopius_ID_version_1/).

# Chapter 3

## Results

### 3.1 Codon saturation

The COI gene region displayed little saturation at all codon positions (Table A5 and Fig. A5). This meant that additional codon models were not necessary in the construction of the COI phylogeny.

### 3.2 Phylogenies

#### 3.2.1 12S

Sequence lengths were approximately 413 bp, with mean nucleotide base frequencies of A ( $44.57 \pm 2.47\%$ ), C ( $12.79 \pm 1.74\%$ ), T ( $38.11 \pm 2.42\%$ ), and G ( $4.54 \pm 2.08\%$ ). The optimal model used was GTR + G. The 12S gene region provided well supported clades for all the *Dactylopius* species included in the analysis, and for three of the six lineages within *D. tomentosus* (Fig. 3.1). 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. See Figure A3 for the resulting Bayesian tree produced from sequences aligned according to their secondary structure. The samples collected from an unidentified *Harrisia* sp. in Windhoek, Namibia, did not successfully amplify using this marker. The *D. ceylonicus* clade showed a well-supported split between the South African and Australian samples. The Bayesian and Maximum Likelihood trees showed the same overall topologies, but the posterior probability and bootstrap support values differed quite substantially for some clades. For example, the split between the *D. tomentosus* ‘californica’ and ‘echinocarpa x acanthocarapa’/‘bigelovii’/‘cylindropuntia’ lineages showed a posterior probability support of 0.8, but

a bootstrap support value of only 47. Similarly, the split between *D. confusus*, and *D. austrinus* and *D. ceylonicus* had a posterior probability support value of 0.9, but a bootstrap support value of only 18. Apart from the outgroup, the greatest within-group p-distances were *D. tomentosus* ‘cholla’ (0.068) and *D. ceylonicus* (0.05), followed by *D. confusus* (0.015) and *D. opuntiae* (0.0005) (Table 3.1). Excluding the outgroup, the average between-group K2P p-distance was  $0.30 \pm 0.14$ .

The haplotype network for *D. opuntiae* (Fig. 3.1) had a total of only seven segregating sites between seven haplotype groups, in contrast to the 16 and 60 found for *D. confusus* and *D. tomentosus*, respectively. The haplotype shared by most individuals comprised of samples from five different groups (MRF, ‘ficus’, ‘stricta’, USA, and Australia), which indicates the absence of intra-specific resolution within *D. opuntiae* using this gene region.

### 3.2.2 18S

Sequence lengths were approximately 570 bp, with mean nucleotide base frequencies of A ( $25.02 \pm 0.60\%$ ), C ( $24.17 \pm 0.61\%$ ), T ( $22.95 \pm 0.74\%$ ), and G ( $27.85\% \pm 0.40$ ). The optimal model used was TIM2ef + I. The 18S gene region showed well-supported clades for all species, except for *Dactylopius austrinus* and *D. ceylonicus*, which grouped into one clade (Fig. 3.2). Within group p-distances for the ingroups were all zero apart from the *D. austrinus* and *D. ceylonicus* group (0.006), and *D. tomentosus* (0.001) (Table 3.2). Excluding the outgroup, the average between-group K2P p-distance was  $0.03 \pm 0.02$ . This marker only showed the *D. tomentosus* ‘cholla’ lineage forming a separate clade, and the ‘echinocarpa x acanthocarpa’ and ‘bigelovii’ grouping together, but did not distinguish between the other three lineages (namely ‘imbricata’, ‘californica’, and ‘cylindropuntia sp.’). It also did not show any intraspecific variation within the *D. opuntiae* clade.

### 3.2.3 COI

Sequence lengths were approximately 603 bp, with mean nucleotide base frequencies of A ( $36.33 \pm 1.61\%$ ), C ( $19.64 \pm 1.66\%$ ), T ( $37.55 \pm 1.19\%$ ), and G ( $6.49 \pm 0.55\%$ ). The optimal model used was HKY + G. COI-A primers amplified only *D. confertus*, *D. opuntiae*, and *D. confusus*, and COI-B primers amplified only *D. tomentosus*. The Bayesian and Maximum Likelihood phylogenies showed the same overall topology, and had well-supported clades for these species, and for the *D. tomentosus* ‘cholla’, ‘imbricata’, and ‘californica’ lineages (Fig. 3.3). Corroborating the 12S phylogeny (Fig. 3.1), there is support for the

‘echinocarpa x acanthocarpa’, ‘bigelovii’, and ‘cylindropuntia sp.’ lineages forming one clade. ‘Bigelovii’ samples differed from ‘echinocarpa x acanthocarpa’ by only one nucleotide. ‘Cylindropuntia sp.’ samples differed from ‘bigelovii’ and ‘echinocarpa x acanthocarpa’ by 4, and 5 nucleotides, respectively. The highest within-group p-distance for the ingroup was for *D. confusus* (0.017), followed by *D. opuntiae* (0.014) and *D. confertus* (0.008) (Table 3.3). Excluding the outgroup, the average between-group K2P p-distance was  $0.22 \pm 0.1$ . The *D. opuntiae* haplotype network contained more haplotypes than represented by the 12S region, with a total of 19 segregating sites (compared to 30 for *D. confusus* and 126 for *D. tomentosus* for COI). Despite this, it also showed that there was not enough genetic variation to tell the ‘ficus’ and ‘stricta’ lineages apart, or to distinguish between different populations.

### 3.2.4 Concatenation of the 12S and 18S gene regions

The 12S and 18S data sets were concatenated due to a significant  $I_{cong}$  value ( $I_{cong} = 1.72$ ,  $p = 2.47 \times 10^{-8}$ ), 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 (Fig. 3.4).

## 3.3 Sample identification

None of the gene phylogenies (12S, 18S, or COI) differentiated between the *D. opuntiae* ‘ficus’ and ‘stricta’ lineages. This was where ISSR analyses were used to gain higher resolution results; which are reported in Section 3.5.

### 3.3.1 South Africa

Using the 12S marker, the unidentified *D. tomentosus* samples collected from *Cylindropuntia fulgida* var. *mammilata* in Jansenville in the Eastern Cape (Extraction IDs: Ekh4, Ekh5, and Ekh6) were confirmed to belong to the ‘cholla’ lineage (Fig. 3.1).

### 3.3.2 Namibia

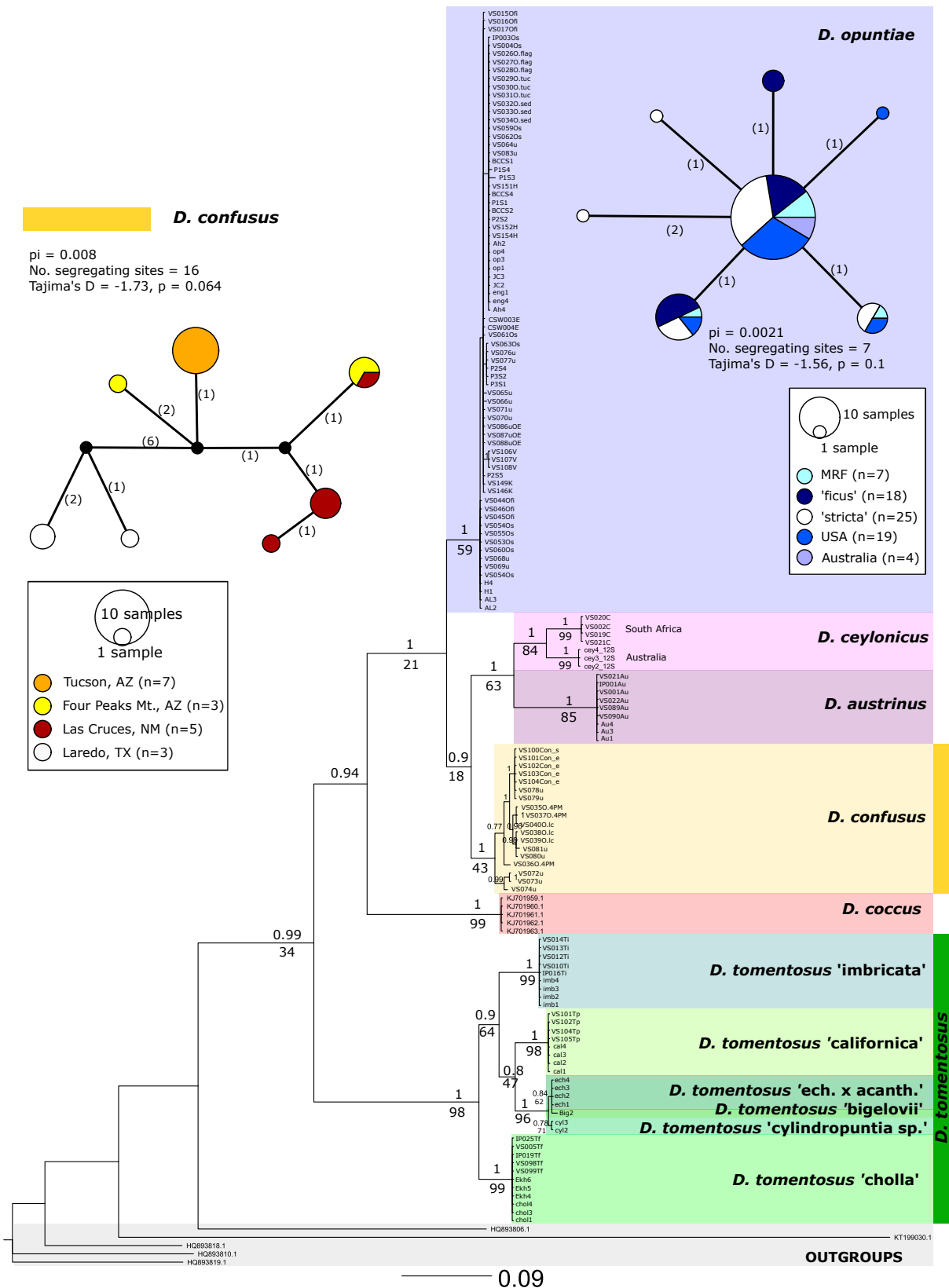
The *Dactylopius* specimens collected on the unidentified *Harrisia* sp. (Extraction IDs: CSW001H and CSW002H) formed a completely separate clade in the 18S and COI phylogenies (Fig. 3.2 and Fig. 3.3). These specimens were sent to Ian Millar, a Hemipteran taxonomist at the Agricultural Research Council (ARC) in Pretoria, who identified them as *D. confertus*. This is the first record of *D. confertus* in Africa.

### 3.3.3 The USA

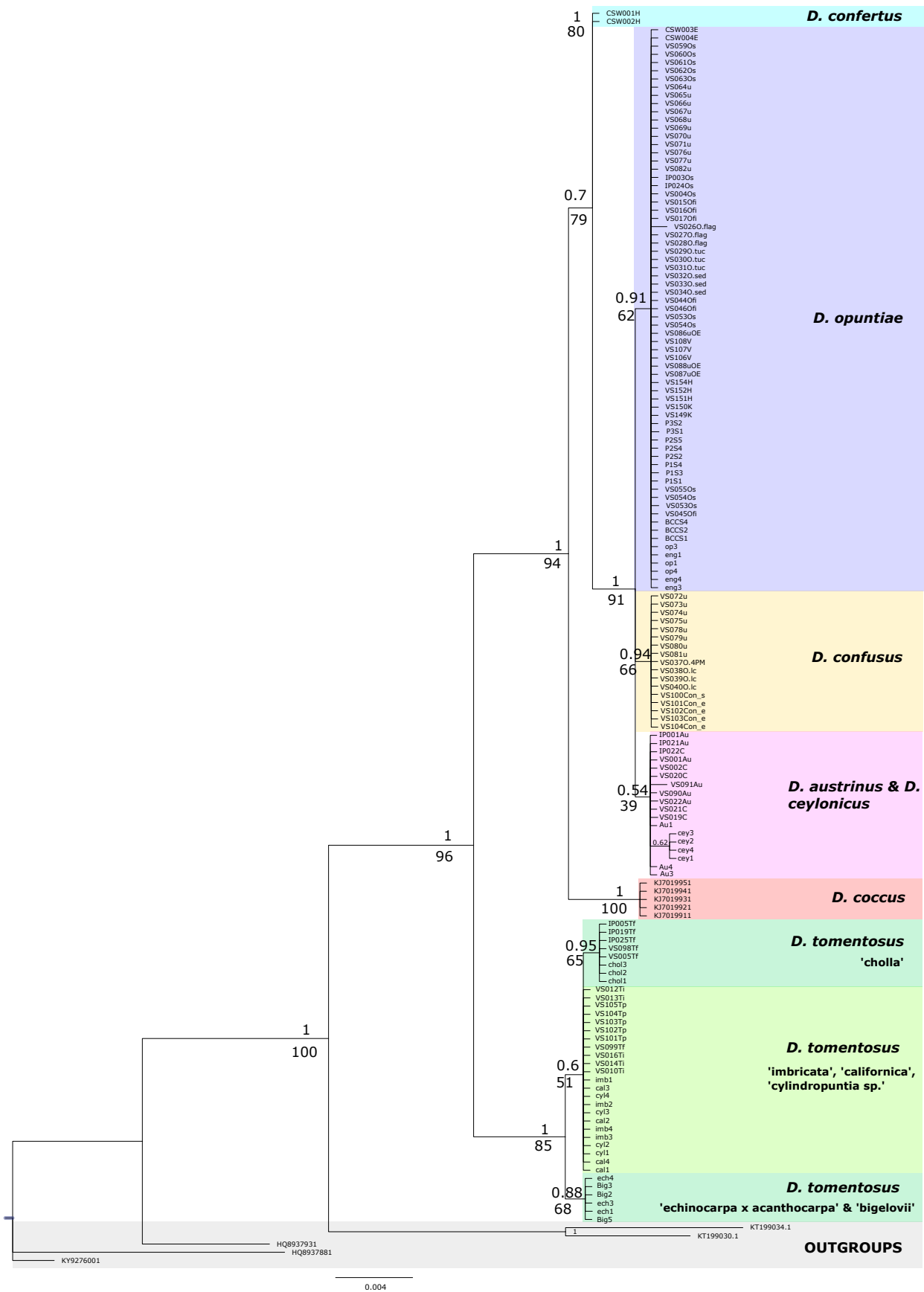
Of the samples collected on *O. engelmannii* in the native range, which were all believed to be *D. opuntiae*, five grouped within the *D. confusus* clade in the 12S (Fig. 3.1), 18S (Fig. 3.2), and COI phylogenies (Fig. 3.3). These were collected in Tucson; Arizona, Laredo; Texas, Las Cruces; New Mexico, and Four Peaks Mountain; Arizona (Extraction IDs: VS035O.4PM, VS036O.4PM, VS037O.4PM, VS038LC, VS039LC, VS040LC, VS072u, VS073u, VS074u, VS078u, VS079u, VS080u, VS081u). Representatives of these five samples were re-identified by Ian Millar, and were confirmed to be *D. confusus*. The original misidentification is because current morphological keys for the Dactylopiidae tend to be inconsistent and subjective. The remaining lineages grouped with *D. opuntiae*, including the ‘Flagstaff’, ‘Sedona’, and ‘Tucson’ lineages (extraction IDs with the suffixes ‘O.flag’, ‘O.sed’, and ‘O.tuc’). Rhodes University Accession numbers for these samples are RH1286 - RH1295 (Table A4).

### 3.3.4 Australia

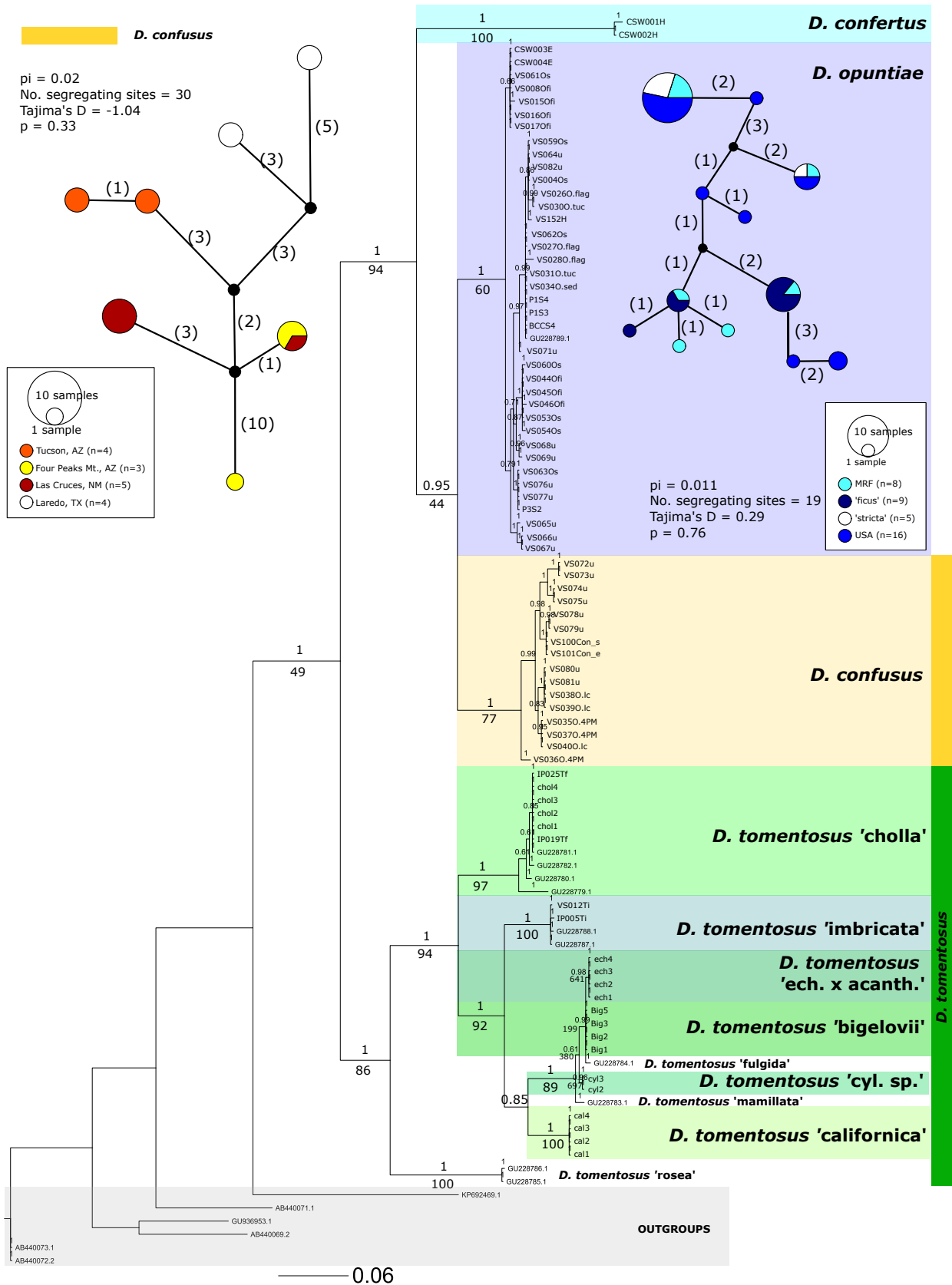
The 12S, 18S, and COI-B primers amplified *Dactylopius tomentosus* samples very well, but only the 12S and COI-B primers clearly differentiated between the ‘imbricata’, ‘californica’, and ‘cholla’ lineages (Fig. 3.1 and Fig. 3.3). The 12S and COI phylogenies grouped the *D. tomentosus* ‘echinocarpa x acanthocarpa’, ‘bigelovii’ and ‘cylindropuntia’ lineages into one clade. The 18S phylogeny showed the grouping of *D. tomentosus* ‘cholla’ into a separate clade, and the ‘echinocarpa x acanthocarpa’ and ‘bigelovii’ lineages into another; while the ‘imbricata’, ‘californica’, and ‘cylindropuntia sp.’ lineages formed polytomies (Fig. 3.2).



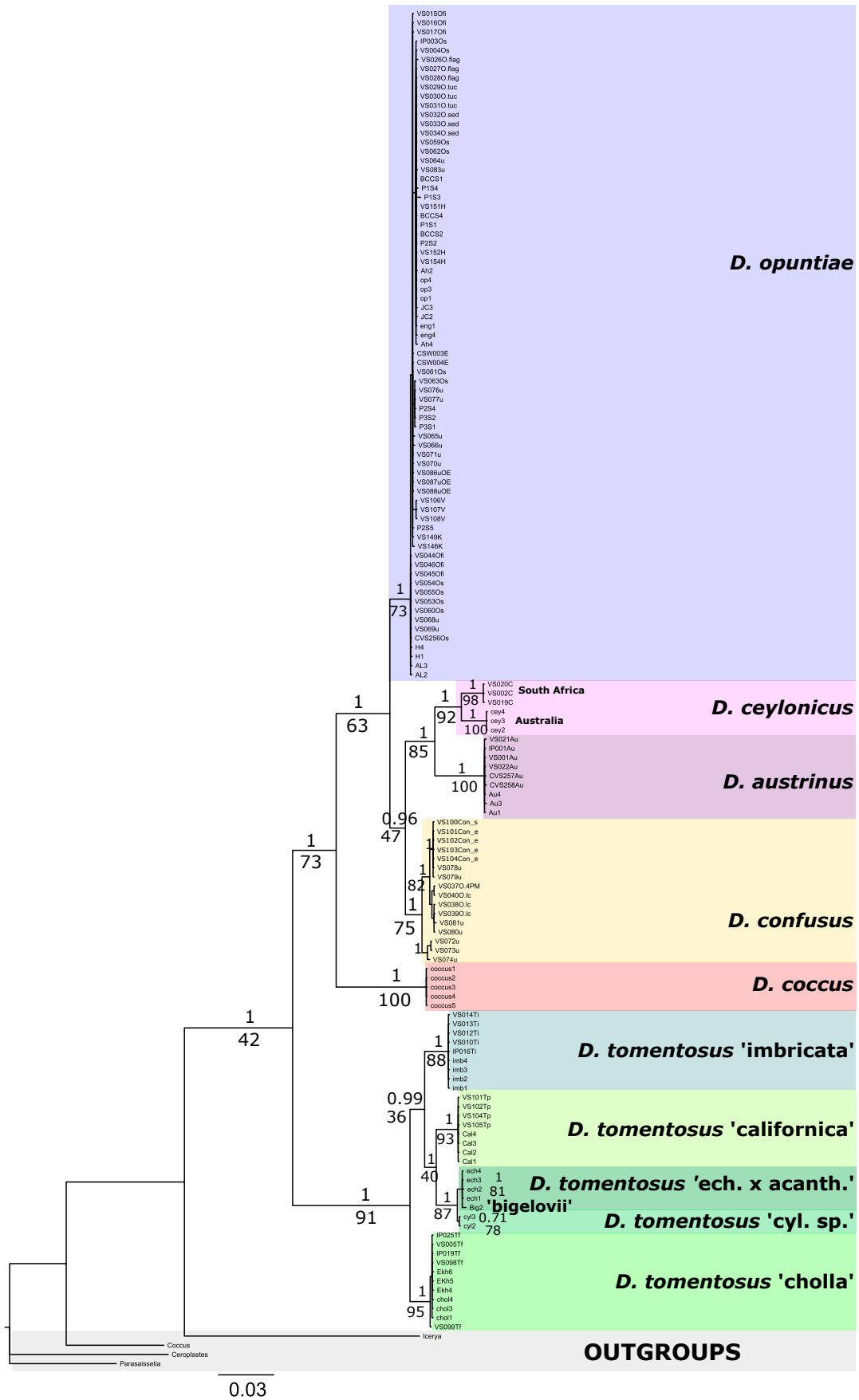
**Figure 3.1:** Bayesian and Maximum likelihood phylogenies for the 12S gene region. Posterior probability values are above, and bootstrap values are below the branches. Haplotype networks are shown inset for *D. opuntiae* and *D. confusus*. Small black dots indicate missing haplotypes, and the diameter of each circle is proportional to the number of genetic sequences that comprise it. Numbers in brackets indicate the number of nucleotide differences between haplotypes.



**Figure 3.2:** Bayesian and Maximum likelihood phylogenies for the 18S gene region. Posterior probability values are above, and bootstrap values are below the branches.



**Figure 3.3:** Bayesian and Maximum likelihood phylogenies for the COI gene region. Posterior probabilities are above, and bootstrap values are below the branches. Haplotype networks are shown for *D. opuntiae* and *D. confusus*. Small black dots indicate missing haplotypes, and the diameter of each circle is proportional to the number of individuals it comprises. Numbers in brackets indicate the number of nucleotide differences between haplotypes.



**Figure 3.4:** Concatenated Bayesian and Maximum Likelihood phylogenies for the 12S and 18S gene regions. Bayesian posterior probability values are above, and Maximum likelihood bootstrap values are below the branches.

**Table 3.1:** Within and between group p-distance matrix for the 12S gene, using the Kimura 2-parameter model. Species are represented by a total of 151 nucleotide sequences consisting of 779 positions. Bold values along the diagonal represent within-group p-distances.

	<i>D. austrinus</i>	<i>D. ceylonicus</i>	<i>D. coccus</i>	<i>D. confusus</i>	<i>D. opuntiae</i>	<i>D. tomentosus</i> 'californica'	<i>D. tomentosus</i> 'cholla'	<i>D. tomentosus</i> 'cylindropuntia sp.'	<i>D. tomentosus</i> 'echinocarpa'	<i>D. tomentosus</i> 'imbricata'	Outgroup
<i>D. austrinus</i> (n = 9)	<b>0.000</b>										
<i>D. ceylonicus</i> n = 6	0.17	<b>0.050</b>									
<i>D. coccus</i> n = 5	0.38	0.33	<b>0.000</b>								
<i>D. confusus</i> n = 18	0.17	0.16	0.30	<b>0.015</b>							
<i>D. opuntiae</i> n = 73	0.18	0.17	0.27	0.11	<b>0.005</b>						
<i>D. tomentosus</i> 'californica' n = 8	0.42	0.39	0.41	0.39	0.42	<b>0.000</b>					
<i>D. tomentosus</i> 'cholla' n = 12	0.39	0.38	0.39	0.37	0.39	0.13	<b>0.068</b>				
<i>D. tomentosus</i> 'cylindropuntia sp.' n = 2	0.42	0.41	0.41	0.40	0.40	0.09	0.14	<b>0.000</b>			
<i>D. tomentosus</i> 'echinocarpa' n = 4	0.42	0.42	0.40	0.40	0.39	0.09	0.15	0.00	<b>0.00</b>		
<i>D. tomentosus</i> 'imbricata' n = 9	0.44	0.40	0.40	0.41	0.42	0.10	0.13	0.09	0.09	<b>0.00</b>	
Outgroup n = 5	0.70	0.66	0.71	0.65	0.67	0.73	0.75	0.73	0.71	0.74	<b>0.903</b>

**Table 3.2:** Within and between group p-distance matrix for the 18S gene, using the Kimura 2-parameter model. Species are represented by a total of 153 nucleotide sequences consisting of 630 positions. Bold values along the diagonal represent within-group p-distances.

	<i>D. austrinus</i> & <i>D. ceylonicus</i>	<i>D. coccus</i>	<i>D. confertus</i>	<i>D. confusus</i>	<i>D. opuntiae</i>	<i>D. tomentosus</i>	<i>D. tomentosus</i> 'echinocarpa' & <i>D. tomentosus</i> 'bigelovii'	<i>D. tomentosus</i> 'cholla'	Outgroup
<i>D. austrinus</i> & <i>D. ceylonicus</i> (n=18)	<b>0.006</b>								
<i>D. coccus</i> (n=5)	0.033	<b>0</b>							
<i>D. confertus</i> (n=2)	0.016	0.020	<b>0</b>						
<i>D. confusus</i> (n=17)	0.010	0.027	0.012	<b>0</b>					
<i>D. opuntiae</i> (n=69)	0.010	0.028	0.013	0.004	<b>0</b>				
<i>D. tomentosus</i> (n=31)	0.057	0.048	0.043	0.047	0.049	<b>0</b>			
<i>D. tomentosus</i> 'echinocarpa' & <i>D. tomentosus</i> 'bigelovii' (n=6)	0.058	0.049	0.043	0.047	0.050	0.004	<b>0</b>		
<i>D. tomentosus</i> 'cholla' (n=5)	0.059	0.047	0.045	0.049	0.051	0.002	0.005	<b>0</b>	
Outgroup (n=5)	0.134	0.129	0.123	0.125	0.127	0.114	0.117	0.116	<b>0.114</b>

**Table 3.3:** Within and between group p-distance matrix for the COI-A and COI-B regions, using the Kimura 2-parameter model. Species are represented by a total of 95 nucleotide sequences consisting of 603 positions. Bold values along the diagonal represent within-group p-distances.

	<i>D. confusus</i>	<i>D. confertus</i>	<i>D. opuntiae</i>	<i>D. tomentosus</i> 'bigelovii'	<i>D. tomentosus</i> 'californica'	<i>D. tomentosus</i> 'cholla'	<i>D. tomentosus</i> 'cylindropuntia sp.'	<i>D. tomentosus</i> 'echinocarpa'	<i>D. tomentosus</i> 'imbricata'	<i>D. tomentosus</i> 'rosea'	Outgroups
<i>D. confusus</i> (n=16)	<b>0.017</b>										
<i>D. confertus</i> (n=2)	0.29	<b>0.008</b>									
<i>D. opuntiae</i> (n=39)	0.13	0.28	<b>0.014</b>								
<i>D. tomentosus</i> 'bigelovii' (n=4)	0.24	0.36	0.23	<b>0.000</b>							
<i>D. tomentosus</i> 'californica' (n=4)	0.24	0.31	0.24	0.09	<b>0.000</b>						
<i>D. tomentosus</i> 'cholla' (n=9)	0.27	0.36	0.25	0.14	0.14	<b>0.009</b>					
<i>D. tomentosus</i> 'cylindropuntia sp.' (n=2)	0.24	0.36	0.24	0.01	0.09	0.15	<b>0.000</b>				
<i>D. tomentosus</i> 'echinocarpa' (n=4)	0.25	0.37	0.23	0.00	0.10	0.15	0.01	<b>0.000</b>			
<i>D. tomentosus</i> 'imbricata' (n=4)	0.26	0.34	0.23	0.10	0.11	0.14	0.10	0.10	<b>0.002</b>		
<i>D. tomentosus</i> 'rosea' (n=2)	0.25	0.31	0.25	0.22	0.23	0.21	0.23	0.22	0.21	<b>0.002</b>	
Outgroups	0.34	0.44	0.32	0.34	0.33	0.32	0.34	0.34	0.32	0.31	<b>0.235</b>

## 3.4 Identification accuracy

### 3.4.1 Nearest Neighbour (NN) results

For species-level identification at a default distance threshold of 1%, the NN algorithm can be a useful test of accuracy. However, when different genes are being analysed that display different evolutionary rates, an adjustable threshold value is required. The NN algorithm tends to result in an abnormally high number of false positive identifications, and is therefore not considered further. For example, Table 3.4 shows a NN ‘True’ outcome of 88.46% for the correct identification of the *D. opuntiae* lineages using the COI region, while the more stringent TID test only had an accuracy rate of 59.62% at a lower threshold of 0.8%.

### 3.4.2 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 3.4). At the lineage level, *Dactylopius 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.

### 3.4.3 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 3.4). 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 intraspecific lineages of *Dactylopius opuntiae* produced 100% ambiguous results for the BCM and TID tests at the 0.1% threshold. 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 Figure 3.2). 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.

### 3.4.4 COI

Identification accuracy at the species level at a 1% threshold was 100% for both the BCM and TID tests (Table 3.4). 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 3.4). 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.

**Table 3.4:** Results of the Best Close Match (BCM), Threshold ID (TID) and Nearest Neighbour (NN) 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.

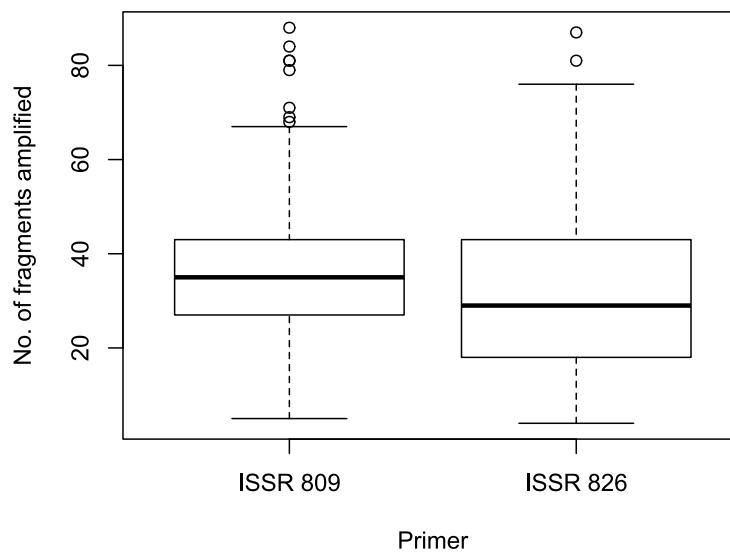
		BCM	Proportion of samples (%)	TID (1%)	Proportion of samples (%)	BCM (optimum %)	Proportion of samples (%)	TID (optimum %)	Proportion of samples (%)	NN	Proportion of samples (%)
<b>12S</b>											
<b>SPECIES LEVEL</b>	Correct	147	100	147	100					<b>T</b>	147 100
										<b>F</b>	0 0
<b>LINEAGE LEVEL</b>	<i>D. opuntiae</i>					<b>0.2%</b>		<b>0.2%</b>			
	Correct	7	15.22	0	0	6	13.04	6	13.04	<b>T</b>	37 80.43
	Incorrect	1	2.17	0	0	1	2.17	1	2.17	<b>F</b>	9 19.57
	Ambiguous	38	82.61	46	100	37	80.43	37	80.43		
	No id	0	0	0	0	2	4.35	2	4.35		
	<i>D. tomentosus</i>							<b>0.2%</b>			
	Correct	34	100.0	28	82.4	34	100.0	34	100.0	<b>T</b>	34 100.0
	Incorrect	0	0.0	0	0.0	0	0.0	0	0.0	<b>F</b>	0 0
	Ambiguous	0	0.0	6	17.6	0	0.0	0	0.0		
<b>18S</b>											
						<b>0.2%</b>		<b>0.2%</b>			
<b>SPECIES LEVEL</b>	Correct	140	94.59	44	29.73	140	94.59	140	94.59	<b>T</b>	148 100
	Ambiguous	8	5.41	104	70.27	8	5.41	8	5.41	<b>F</b>	0 0
<b>LINEAGE LEVEL</b>	<i>D. opuntiae</i>							<b>0.1%</b>			
	Correct			0	0			0	0		
	Incorrect			0	0			0	0		
	Ambiguous			35	100			35	100		
	No id			0	0			0	0		
	<i>D. tomentosus</i>					<b>0.1%</b>		<b>0.1%</b>			
	Correct	8	22.22	0	0	8	22.22	8	22.22	<b>T</b>	8 22.22
	Incorrect	0	0	0	0	0	0	0	0	<b>F</b>	28 77.78
	Ambiguous	28	77.78	36	100	28	77.78	28	77.78		
<b>COI</b>											
<b>SPECIES LEVEL</b>	Correct	83	100	83	100					<b>T</b>	83 100
										<b>F</b>	0 0
<b>LINEAGE LEVEL</b>	<i>D. opuntiae</i>					<b>0.8%</b>		<b>0.8%</b>			
	Correct	32	61.54	24	46.15	32	61.54	31	59.62	<b>T</b>	46 88.46
	Incorrect	3	5.77	1	1.92	3	5.77	1	1.92	<b>F</b>	6 11.54
	Ambiguous	17	32.69	27	51.92	17	32.69	20	38.46		
	<i>D. tomentosus</i>					<b>0.2%</b>		<b>0.2%</b>			
	Correct	26	96.3	16	59.26	24	88.89	19	70.37	<b>T</b>	27 100
	Incorrect	0	0	0	0	0	0	0	0	<b>F</b>	0 0
	Ambiguous	0	0	10	37.04	0	0	5	18.52		
	No ID	1	3.7	1	3.7	3	11.11	3	11.11		
	<i>D. tomentosus</i> (G)					<b>3.3%</b>		<b>3.3%</b>			
	Correct	26	96.3	26	96.3	27	100	27	100	<b>T</b>	27 100
	Incorrect	0	0	0	0	0	0	0	0	<b>F</b>	0 0
	Ambiguous	0	0	9	0	0	0	0	0		
	No ID	1	3.7	1	3.7	0	0	0	0		

### 3.5 ISSR fragment analysis

ISSR 809 had more peaks, and lower Jaccard and Euclidean error rates than ISSR 826 (Table 3.5 and Fig. 3.5). See Figures A7 and A8 for additional statistics regarding the error rates for both ISSR primers, and Tables A6 and A7 and Figure A12 for statistics regarding measures of genetic diversity.

**Table 3.5:** 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 $\pm$ sd	Max. peaks	Min. peaks	Avg. Euclidean error rate $\pm$ sd (Bonin <i>et al.</i> , 2004)	Avg. Jaccard error rate $\pm$ sd (Holland <i>et al.</i> , 2008)
ISSR 809	384	36.58 $\pm$ 15.75	88	5	0.08 $\pm$ 0.03	0.48 $\pm$ 0.14
ISSR 826	366	32.09 $\pm$ 18.93	87	4	0.09 $\pm$ 0.03	0.53 $\pm$ 0.17

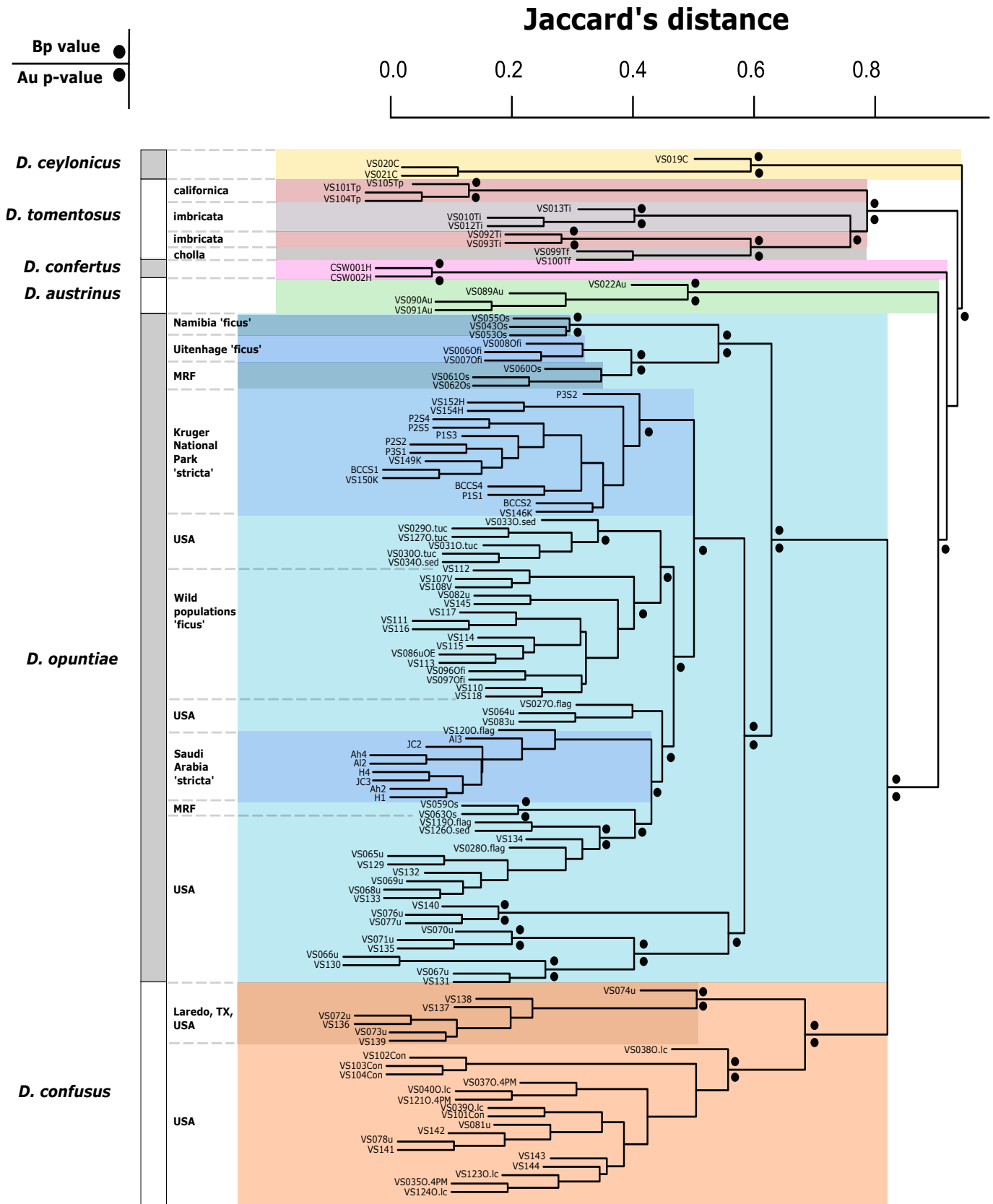


**Figure 3.5:** A box plot for the number of fragments amplified by the ISSR 809 and ISSR 826 primers.

#### 3.5.1 Hierarchical clustering tree

The UPGMA clustering tree representative of the combination of both ISSR primers showed a separation of all the *Dactylopius* species (Fig. 3.6), and corroborated the topologies in the 12S, 18S, and COI

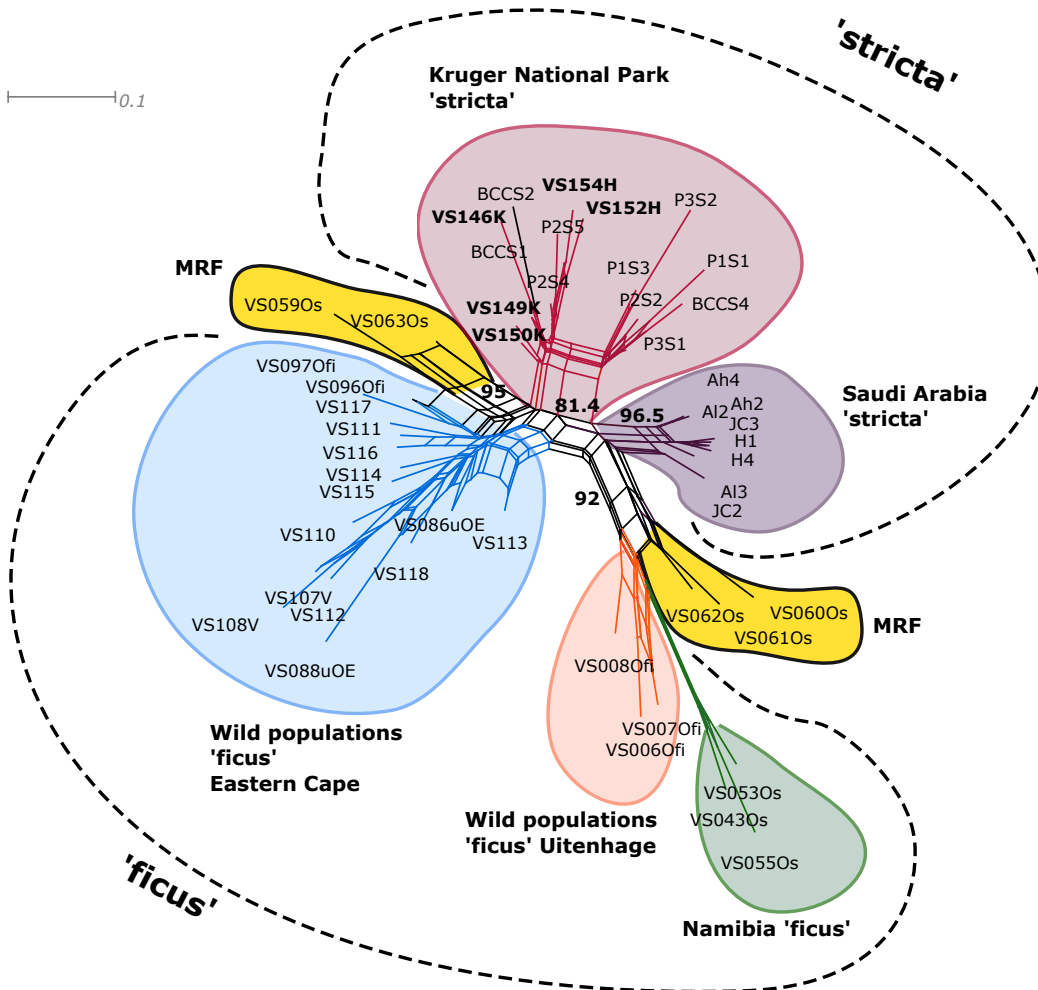
phylogenies. *Dactylopius tomentosus* separated into the ‘californica’, ‘imbricata’, and ‘cholla’ lineages, but ‘imbricata’ samples split into two separate groups. The Uitenhage ‘ficus’ and Mass Rearing Facility samples grouped together as a sister group to the Namibia ‘ficus’ samples. The ‘stricta’ samples from the Kruger National Park, Saudi Arabia, and samples from ‘stricta’ source populations (kept by John Hoffmann and Hildegard Klein) separated from the ‘ficus’ samples and from those collected in the United States. Samples from the Pecos County in Texas in the USA (VS082u and VS145) grouped with the wild ‘ficus’ samples collected in the Eastern Cape, South Africa. Two MRF samples (Extraction IDs: VS059Os and VS063Os) grouped with samples from the USA.



**Figure 3.6:** Hierarchical clustering tree (UPGMA method) applying Jaccard's transformation for the ISSR binary data obtained from two primers (ISSR 809 and ISSR 826). Approximately unbiased p-values (Au p-value) and the bootstrap probability values (Bp-value) above 70 are shown as black dots above and below the relevant clades. One thousand bootstrap repetitions were run. MRF = Uitenhage Mass Rearing Facility.

### 3.5.2 SplitsTree

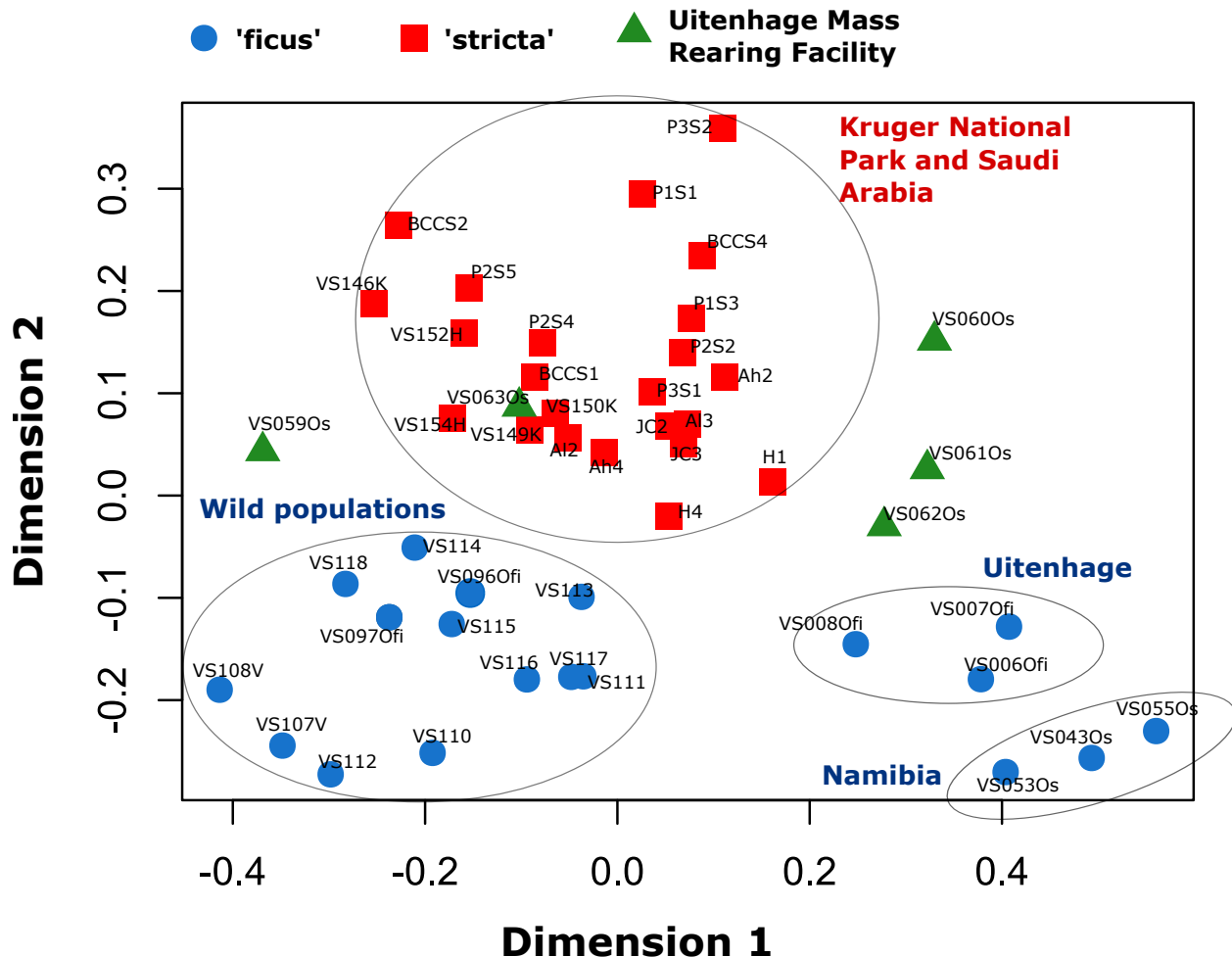
Overall, the ‘stricta’ and ‘ficus’ lineages showed a clear separation (Fig. 3.7). The ANOSIM results in Table 3.6 suggest that the MRF samples were not significantly different to ‘ficus’, but that they were to ‘stricta’. The Kruger National Park and Saudi Arabian ‘stricta’ samples separated into two groups, and the John Hoffman and Hildegard Klein samples fell into the Kruger National Park group. The wild populations, Namibian, and Uitenhage ‘ficus’ samples also formed separate groups. See Figures A9 and A10 for SplitsTree diagrams for each individual ISSR primer.



**Figure 3.7:** SplitsTree graphical output (NeighborNet method, applying Jaccard’s index to calculate the distance matrix) for *Dactylopius opuntiae* individuals (ISSR 809 and ISSR 826). The Uitenhage and Namibia clusters are known to be the ‘ficus’ lineage, and the Kruger National Park (KNP) contain the ‘stricta’ lineage. Individuals in the ‘wild populations’ group were sampled from *Opuntia ficus-indica* and *O. engelmannii* host plants in the Eastern Cape. MRF = the Mass Rearing Facility at Uitenhage (Eastern Cape). Bootstrap values above 75 are shown. Bold names in the KNP group denote the known ‘stricta’ source population samples (kept by Hildegard Klein and John Hoffmann).

### 3.5.3 Non-metric Multi-dimensional Scaling (nMDS) Plot

A non-metric MDS plot (Fig. 3.8) showed a very similar output to the SplitsTree diagram in Figure 3.7, with the difference of one of the Uitenhage MRF samples (VS0630s) falling within the ‘stricta’ group.



**Figure 3.8:** A non-metric MDS plot based on the binary data (applying Jaccard’s distance transformation) obtained from ISSR 809 and ISSR 826 to represent two different *Dactylopius opuntiae* lineages. Blue circles = ‘ficus’, red squares = ‘stricta’, and green triangles = the Uitenhage Mass Rearing Facility.

The  $K = 2$  dimension was applied for the nMDS plot, as the resulting stress value fell within the acceptable range (Fig. A11 A, B and C) (Dugard *et al.*, 2010). With the ‘ficus’ group as a whole (ANOSIM:  $R = 0.43$ ,  $p = 0.001$ ; PERMANOVA:  $DF = 2$ ,  $F = 6.6$ ,  $R^2 = 0.22$ ,  $p = 0.001$ ), and divided into known ficus samples and wild populations (ANOSIM:  $R = 0.72$ ,  $p = 0.001$ ; PERMANOVA:  $DF = 3$ ,  $F = 11.69$ ,  $R^2 = 0.43$ ,  $p = 0.001$ ), pairwise comparisons showed that all groups were significantly different from each other (Tables 3.6, 3.7, 3.8 and 3.9). The ‘ficus’ and MRF groups showed the greatest level of similarity (Tables 3.6 and 3.9).

**Table 3.6:** ANOSIM intra-group statistics for *D. opuntiae* ‘stricta’, ‘ficus’ and Mass Rearing Facility (MRF) groups, showing their R values, followed by uncorrected p-values in brackets. Permutations = 999. Overall R value = 0.44, p = 0.001. Asterisks denote significance (p < 0.05).

	‘stricta’ (n = 23)	‘ficus’ (n = 21)	MRF (n = 5)
‘stricta’			
<b>ficus</b>	0.46 (0.0001)*		
<b>MRF</b>	0.57 (0.0001)*	0.20 (0.066)	

**Table 3.7:** ANOSIM intra-group statistics for *D. opuntiae*, where the ‘ficus’ group is split into wild populations and the ‘ficus’ collected from around Uitenhage and in Namibia. R values are shown, followed by uncorrected p-values in brackets. Permutations = 999. Overall R value = 0.72, p = 0.001. Asterisks denote significance (p < 0.05).

	‘stricta’ (n = 23)	‘ficus’ (n = 6)	MRF (n = 5)	Wild (n = 15)
‘stricta’				
<b>‘ficus’</b>	0.90 (0.0001)*			
<b>MRF</b>	0.57 (0.0001)*	0.43 (0.0043)*		
<b>Wild</b>	0.66 (0.0001)*	0.96 (0.0001)*	0.74 (0.0002)*	

**Table 3.8:** Permutational Multivariate Analysis Of Variance Using Distance Matrices for *D. opuntiae* ‘stricta’, ‘ficus’ and Mass Rearing Facility (MRF) groups. Pairwise p-values are shown. Permutations = 999, DF = 2, F = 6.60, R<sup>2</sup> = 0.22, p = 0.001. Asterisks denote significance (p < 0.05).

	‘stricta’	‘ficus’
<b>‘ficus’</b>	0.003*	
<b>MRF</b>	0.003*	0.045*

**Table 3.9:** Permutational Multivariate Analysis Of Variance Using Distance Matrices intra-group statistics for *D. opuntiae* where the ‘ficus’ group is split into wild populations and the ‘ficus’ collected from around Uitenhage and in Namibia. MRF = Mass Rearing Facility. Pairwise p-values are shown. Permutations = 999, DF = 3, F = 11.69, R<sup>2</sup> = 0.43, p = 0.001. Asterisks denote significance (p < 0.05).

	‘stricta’	ficus	MRF
<b>‘ficus’</b>	0.0012*		
<b>MRF</b>	0.0012*	0.006*	
<b>Wild</b>	0.0012*	0.0012*	0.0012*

### 3.5.4 Structure plots

With population data not set as priors, the Puechmaille (2016) method indicated an optimal  $K$ -value of 5, set at a threshold of 0.6, for all four tests (the MedMedK, MedMeanK, MaxMedK and MaxMeanK). The Evanno *et al.* (2005) method indicated optimal  $K$ -values of 2 and 6 according to the  $\Delta K$  and mean LnP( $K$ ) methods, respectively.

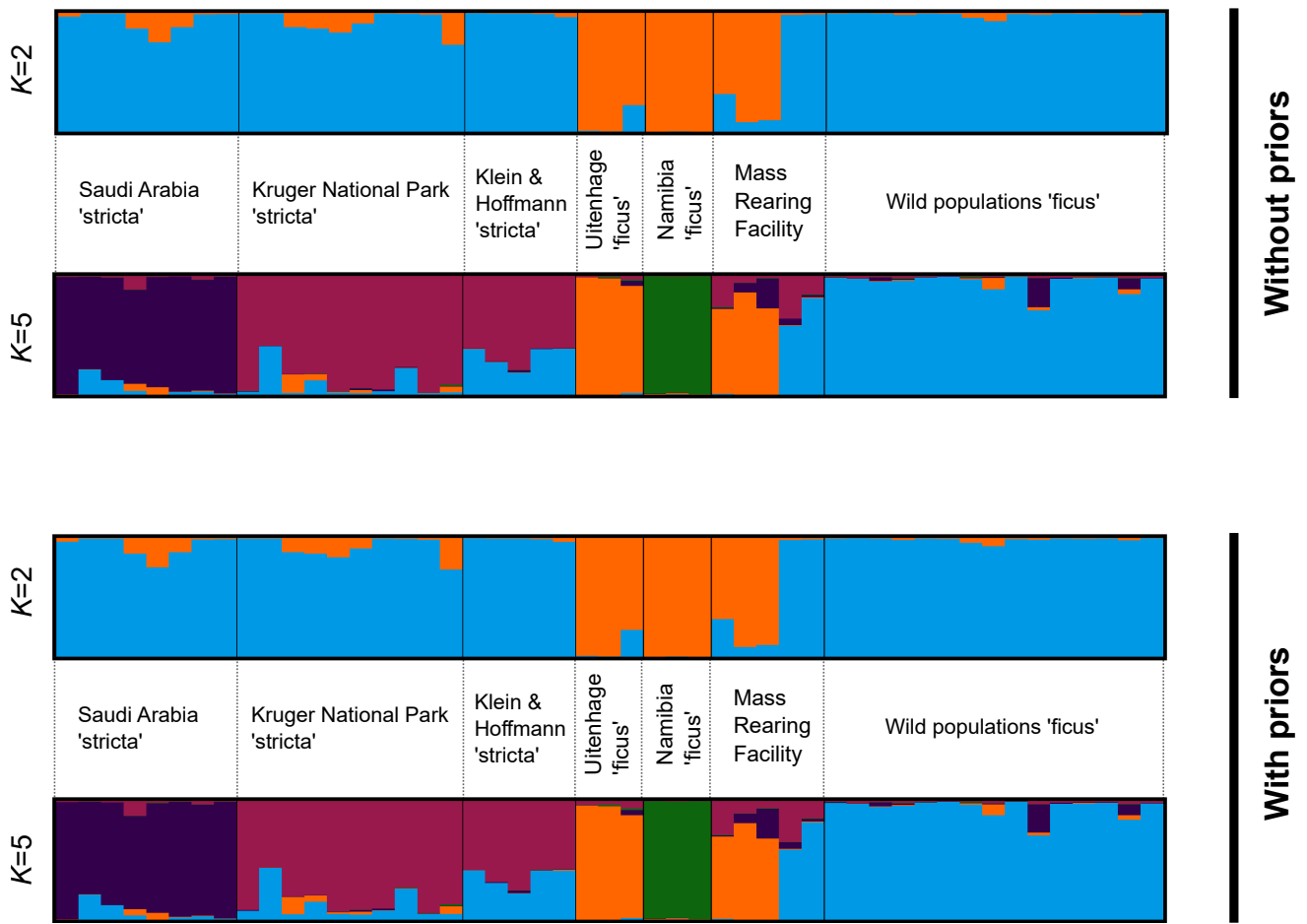
When population data were set as priors, the Puechmaille (2016) method gave the same output as above, except that the Evanno *et al.* (2005) method gave the optimal  $K$ -value as being 2 for the  $\Delta K$  method, and 5 for the mean  $\text{LnP}(K)$  method. Clusters  $K = 2$  and  $K = 5$  are presented in the resulting Structure plots (Fig. 3.9). In both Structure plots (with and without priors), when  $K = 2$ , the Saudi Arabia, Kruger National Park and Klein and Hoffmann source populations of ‘stricta’ formed one shared cluster (light blue). The Uitenhage and Namibia ‘ficus’ and three of the MRF samples formed a second cluster (orange). Two of the MRF samples shared the same cluster group as ‘stricta’. When forced into one of two clusters, the wild ‘ficus’ populations were more genetically similar to the ‘stricta’ cluster than they were to ‘ficus’ samples. When  $K = 5$ , the Kruger National Park and Klein and Hoffmann source population ‘stricta’ samples formed one cluster (maroon) (Fig. 3.9). The Uitenhage ‘ficus’ and three of the MRF samples formed a separate cluster (orange). The remaining two MRF samples shared a cluster with the wild ‘ficus’ populations (light blue). The MRF samples shared some similarity to the Kruger National Park and Saudi Arabian ‘stricta’ clusters. The Saudi Arabian ‘stricta’ (dark purple) and Namibian ‘ficus’ (green) formed their own separate clusters. Taking into account Figures 3.7, 3.8, and 3.9, individuals from the Uitenhage MRF appear to be either hybrids of the ‘stricta’ and ‘ficus’ lineages, or they fall within the ‘ficus’ group.

### 3.5.5 Identification accuracy

At the species level, at an optimal threshold of 60%, all barcode tests (BCM, TID, and NN) had a 100% IA (Table 3.10). At the lineage level, at an optimal threshold of 45%, *D. tomentosus* had an IA of 100% for both the BCM and TID tests (Table 3.10). 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.

**Table 3.10:** Barcode testing results for the Best Close Match (BCM), Threshold ID (TID), and Nearest Neighbour (NN) tests, at the species and lineage level.

		BCM	% of samples	TID	% of samples	NN	% of samples
<b>SPECIES LEVEL</b>		<b>60% threshold</b>		<b>60% threshold</b>			
	Correct	124	100	124	100	True	124 100
	No id	0	0	0	0	False	0 0
<b>LINEAGE LEVEL</b>							
	<i>D. tomentosus</i>	<b>45% threshold</b>		<b>45% threshold</b>			
	Correct	10	100	10	100	True	10 100
	Incorrect	0	0	0	0	False	0 0
	Ambiguous	0	0	0	0		
	<i>D. opuntiae</i>	<b>45% threshold</b>		<b>45% threshold</b>			
	Correct	36	81.82	36	81.82	True	44 100
	Incorrect	0	0	0	0	False	0 0
	Ambiguous	0	0	0	0		
	No id	8	18.18	8	18.18		



**Figure 3.9:** STRUCTURE results using Bayesian analysis (MCMC = 50 000, burn-in = 100 000), showing genetic clustering patterns for *Dactylopius opuntiae* individuals. Results for both  $K = 2$  and  $K = 5$  are shown for runs with and without priors.

# Chapter 4

## Discussion

This chapter summarises and explains the results obtained from two mitochondrial regions, one nuclear gene region, and two ISSRs used to assess the identification accuracy of the Dactylopiidae using both tree-based methods and barcode testing algorithms. It also discusses aspects relating to geographical distribution, and the practical applications arising from this thesis.

### 4.1 Genetic markers

#### 4.1.1 12S

The 12S rRNA genetic marker provided 100% identification accuracy (IA) at both the species level (at a 1% genetic distance threshold) and for three of the six *D. tomentosus* lineages (at a 0.2% genetic distance threshold) using barcode testing algorithms. Bayesian (BI) and Maximum Likelihood (ML) phylogenies supported this result, showing well-supported clades that distinguished all species. Both standard and secondary structure RNA alignment methods suggest that this marker can accurately distinguish between the ‘californica’, ‘cholla’, and ‘imbricata’ lineages, but that the ‘echinocarpa x acanthocarpa’, ‘bigelovii’, and ‘cylindropuntia spp.’ samples form one clade.

The BI tree showed a well-supported separation between South African and Australian *D. ceylonicus* samples. This is most likely due to the fact that the species was released in South Africa and Australia in 1913 and 1914, respectively, and have been isolated from each other for more than a hundred years. The agents used in these two releases came from the same stock, which were originally collected in Brazil, then exported to India, and subsequently to Sri Lanka (Winston *et al.*, 2014). *Dactlopius austrinus* and *D. tomentosus* ‘imbricata’ and ‘cholla’ did not show the same separation between the two countries, most likely because they have not been separated for long enough.

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 ( $0.30 \pm 0.14$ ). 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 BI tree aligned according to the secondary structure of RNA did not vary greatly from the standard alignment, but did show stronger support for a single clade containing the *D. tomentosus* ‘echinocarpa x acanthocarpa’ and ‘bigelovii’ lineages that was separate from the ‘cylindropuntia sp.’ samples.

A number of other studies have found the 12S rRNA marker highly useful for DNA fingerprinting, such as for the identification of ticks (Acari: Ixodida) (Lv *et al.*, 2014), earthworms (Lumbricidae) (Klarica *et al.*, 2012), commercial fishes (Ardura *et al.*, 2010; Hardy *et al.*, 2011; Cawthorn *et al.*, 2012), snakes (from venom samples) (Pook and McEwing, 2005), and various animal species used in traditional medicines (Luo *et al.*, 2011). Based on the results of this study, and due to the ease of PCR amplification of the Dactylopiidae using this gene region, it is recommended as the most efficient for species-level identification, and for the identification of the ‘cholla’, ‘californica’, and ‘imbricata’ lineages within the *D. tomentosus* species.

#### 4.1.2 18S

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. This may suggest a more recent speciation event between these two species, relative to the others. The overall IA was 94.59% at the species level. The primers successfully amplified the target region of all six *Dactylopius* species used in this study. The genetic distance threshold for this marker was lower (0.2%) for species level identification, and had the lowest mean between-group K2P distance ( $0.03 \pm 0.02$ ) owing to the slow evolutionary rate in this nuclear gene relative to mitochondrial regions. Evans *et al.* (2007), for example, tested COI, rbcL, 18S and ITS rDNA as potential barcoding regions, and found that 18S was the least variable marker.

Of the *D. tomentosus* lineages, only ‘cholla’, and ‘echinocarpa x acanthocarpa’ and ‘bigelovii’ formed separate clades, while ‘imbricata’, ‘californica’, and ‘cylindropuntia’ were unresolved. In their study of Murray-Darling Basin fishes, Hardy *et al.* (2011) found that the 18S rRNA marker could successfully

identify 83.1% of the species sampled, but that there were a number of different species within a genus complex that shared identical sequences. In the study by Campana *et al.* (2015), no intraspecific variation was found within the seven sequences obtained from *D. coccus* populations. This corroborates the findings of the current project, 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, particularly when supplementing a concatenated genetic data set.

### 4.1.3 COI

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 optimal for assessing identification accuracy at the species level, which is consistent with the barcoding literature (Hebert *et al.*, 2003a,b). It is not, however, able to distinguish between *D. opuntiae* lineages. When barcode accuracy was tested at the lineage level, all *D. opuntiae* sequences displayed negative barcode gaps (intraspecific variation > interspecific variation) due to the high similarity between the ‘ficus’ and ‘stricta’ lineages. The average within-group p-distance was similar to that of the 12S region ( $0.22 \pm 0.1$ ).

COI-B primers successfully amplified all *D. tomentosus* lineages. At a genetic distance threshold of 3.3%, IA increased from 70.37% to 100% when the ‘echinocarpa x acanthocarpa’, ‘bigelovii’, and ‘cylindropuntia sp.’ lineages were treated as one lineage. ISSR analyses for these particular lineages may reveal further differences between them.

## 4.2 Distance-based barcode testing

In agreement with Birch *et al.* (2017), the TID barcode testing algorithm is the more conservative test, and is recommended as the most appropriate since it takes into consideration a genetic distance threshold. It, therefore, reduces the number of false positive identifications, as opposed to the NN and BCM tests. The NN algorithm was mostly an inappropriate measure in the context of this study, particularly with reference to the lineage level where genetic similarity was very high. Since each gene evolves at a different rate, a one-size-fits-all approach of a default 1% distance threshold is inaccurate. Combining appropriate distance and tree-based methods to determine barcoding accuracy offers a more thorough approach to

the identification of target taxa.

### 4.3 Posterior probabilities vs bootstrap values

Some clades in the phylogenies in this study showed much higher bayesian posterior probability (BPP) support than their corresponding ML bootstrap values. This is not uncommon, and stems from the fact that these two methods assess data in fundamentally different ways and are based upon different statistical assumptions (Huelsenbeck *et al.*, 2002; Douady *et al.*, 2003; Garcia *et al.*, 2004; Huelsenbeck and Rannala, 2004). Huelsenbeck *et al.* (2002) suggest that the ML bootstrap method underestimates the confidence of phylogenetic clades, and propose that a Bayesian analysis is the most easily interpretable and accurate given the correct evolutionary models. In a simulation, Alfaro *et al.* (2003) found that the greatest disparity between BPP and bootstrap values occurred when the branch lengths were short, or when the number of characters available for comparison was small. They suggest that a Bayesian analysis offers a greater sensitivity to phylogenetic signal, and that a substantial increase in the number of characters will facilitate the convergence of bootstrap values to BPP.

A study by Hardy *et al.* (2011) used the 12S rRNA region (in addition to 18S rRNA and the mtDNA control region (CR)) to differentiate between different fish species in the Murray-Darling River Basin in Australia. The BI and ML analyses did not differ substantially in their support values, but they did report that the 12S trees in particular showed greater variability (where a number of clades had bootstrap support values  $< 75$  and BPP  $> 95$ ). Similarly, a study by Wilcox *et al.* (2002) analysed the 12S, 16S, and valine-tRNA gene regions of dwarf boa snakes (Tropidophiidae), and found that some branches in their 12S and 16S phylogenies had BPP values more than double that of the corresponding bootstrap support values. Based on their results, the authors recommended the use of Bayesian analyses above bootstrapping methods, as it provided a more reliable estimation of phylogenetic accuracy. The present study supports this statement, especially taking into account the faster computational time that a Bayesian analysis offers compared to other methods (Huelsenbeck *et al.*, 2001; Douady *et al.*, 2003).

## 4.4 ISSR fragment analysis

The percentage of polymorphism found in this study was lower than that reported by Silva *et al.* (2013), who collected *D. opuntiae* on *O. ficus-indica* host plants in the native range in northeast Brazil. Their overall polymorphism percentage was 70.9%, taken as an average of nine different ISSR primers. The present study reported only 44.93% polymorphism. Additionally, they reported a genetic similarity of 80% across all populations. Comparatively, the present study reported an average Jaccard similarity of 51.67% (SE  $\pm$  1.01) for all *D. opuntiae* samples. This is arguably due to the higher number of fragments reported in the present study due to the use of DNA fragment analysis using capillary electrophoresis. Over three-fold more fragments were found per primer than by Silva *et al.* (2013), who scored fragments manually through visualising bands on agarose gels. Silva *et al.* (2013) used both ISSR and RAPD primers to distinguish between *D. opuntiae* populations, and concluded that the RAPD primers used were more effective than ISSRs in grouping populations, but neither method succeeded in discriminating between populations from different geographic regions. The *D. opuntiae* ISSR data using Bayesian and ordination methods showed a statistically significant difference between the ‘ficus’ and ‘stricta’ lineages that the traditional DNA gene regions failed to reveal. The methods also showed significant differences between *D. opuntiae* ‘ficus’ populations; namely wild populations collected on *O. ficus-indica* and *O. engelmannii* in the Eastern Cape, and specimens collected from *O. ficus-indica* around the Uitenhage Mass Rearing Facility (MRF) and in Namibia.

### 4.4.1 ‘Stricta’ and ‘ficus’ lineages

#### Eastern Cape wild populations

An unexpected result arising from the Bayesian Structure analysis (and supported by the nMDS ordination grouping) was that when forced into one of two cluster groups, the wild ‘ficus’ populations collected from *O. ficus-indica* and *O. engelmannii* in the Eastern Cape were more genetically similar to the ‘stricta’ cluster than they were to known ‘ficus’ samples. It is possible that the wild populations have regularly hybridised with ‘stricta’ and ‘ficus’ lineages (hybridisation between these two lineages is reported in the work by Hoffmann *et al.* (2002) and Hoffmann (2003)), and that the genotypic status detected by the two specific ISSR primers used in this study currently aligns more with that of ‘stricta’. Collecting more

specimens from *O. ficus-indica* and *O. engelmannii* in more distant regions of the country (in addition to the Namibian samples included in this study that grouped in the ‘ficus’ cluster) could provide further clarity. When forced into one of five cluster groups, the wild populations did represent a distinct genotypic cluster not shared by any of the ‘stricta’ genotypes.

### **Uitenhage Mass Rearing Facility (MRF)**

The *D. opuntiae* population believed to be of the pure-bred ‘stricta’ lineage at the MRF were more similar to ‘ficus’ samples collected from *O. ficus-indica* growing around the facility and from Namibian samples than they were to ‘stricta’ samples. The MRF individuals are either the result of crosses between the two lineages, or they are merely part of the ‘ficus’ group. This suggests that contamination may have taken place, and that the control efficacy of the hybrid individuals being reared may be compromised. It is recommended that 1) the current culture be destroyed, and replaced with a new colony that is known to be the ‘stricta’ lineage, 2) the wild ‘ficus’ populations around the MRF should be killed, and 3) the genetic status of the colonies be checked on a regular basis to ensure that a pure-bred lineage is maintained. Additionally, it would be worth crossing and backcrossing the pure ‘stricta’, ‘ficus’, and MRF populations to test their host-specificities.

### **The Kruger National Park and Saudi Arabia**

The *D. opuntiae* samples from the Kruger National Park shared a genotypic cluster with the known ‘stricta’ source populations (kept by Hildegard Klein and John Hoffmann), and were different from the ‘ficus’ clusters. It could therefore be confirmed that these samples were of the ‘stricta’ lineage. Similarly, the Saudi Arabian *D. opuntiae* samples collected from *O. stricta* shared a genetic cluster with the ‘stricta’ source populations; confirming their identity as well. These populations can therefore effectively control *O. stricta* (Githure *et al.*, 1999; Volchansky *et al.*, 1999), and can be mass reared if necessary. These populations can also be used as a source for the new clean culture of ‘stricta’ at the Uitenhage MRF.

## **4.4.2 ISSR identification accuracy**

Due to the different kind of data that ISSR analyses produce, the distance threshold required to create species 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.

The identification success rate for *D. opuntiae* lineages was 81.82% at a 45% distance threshold, with the remaining 18.18% falling into the ‘No ID’ category. 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 (‘imbricata’, ‘cholla’, and ‘californica’) showed a 100% IA at a distance threshold of 45%, although using the 12S rRNA marker to identify these is less labour intensive.

Of the two ISSR primers used, 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. The average Euclidean error rates of 0.08 ( $\pm 0.03$ ) and 0.09 ( $\pm 0.03$ ) for ISSR 809 and ISSR 826, respectively, were in the same range as that found by Paterson and Zachariades (2013) for *Chromolaena odorata*. Whether the shared absence of bands is included in the error rate calculation can greatly impact its interpretation, and so it is important that future studies include both the Euclidean and Jaccard error rates for comparison. The present study reported average Jaccard error rates 6-fold greater than their Euclidean equivalent, which is comparable to those found by Holland *et al.* (2008), who reported a 4-fold difference in one experiment. Although a 81.82% 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)<sub>7</sub> 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.

## 4.5 Geographical distribution

The first phylogeny for the Dactylopiidae was constructed by Pérez-Guerra and Kosztarab (1992) using morphological data. The authors concluded that all the *Dactylopius* species derived from one common ancestor, that *D. coccus* was the most basal species, and that *D. ceylonicus* was the most derived. Rodríguez *et al.* (2001) later created a phylogeny based on both morphological traits and geographical distributions, which formed two clades separating North and South American species. The results from the present study, and those from a molecular analysis by Ramírez-Puebla *et al.* (2010) (using the same 12S and 18S genetic markers), did not find this well-defined separation between the northern and southern continents (Fig. 4.1). Both studies did, however, agree with the finding that *D. tomentosus* was the most

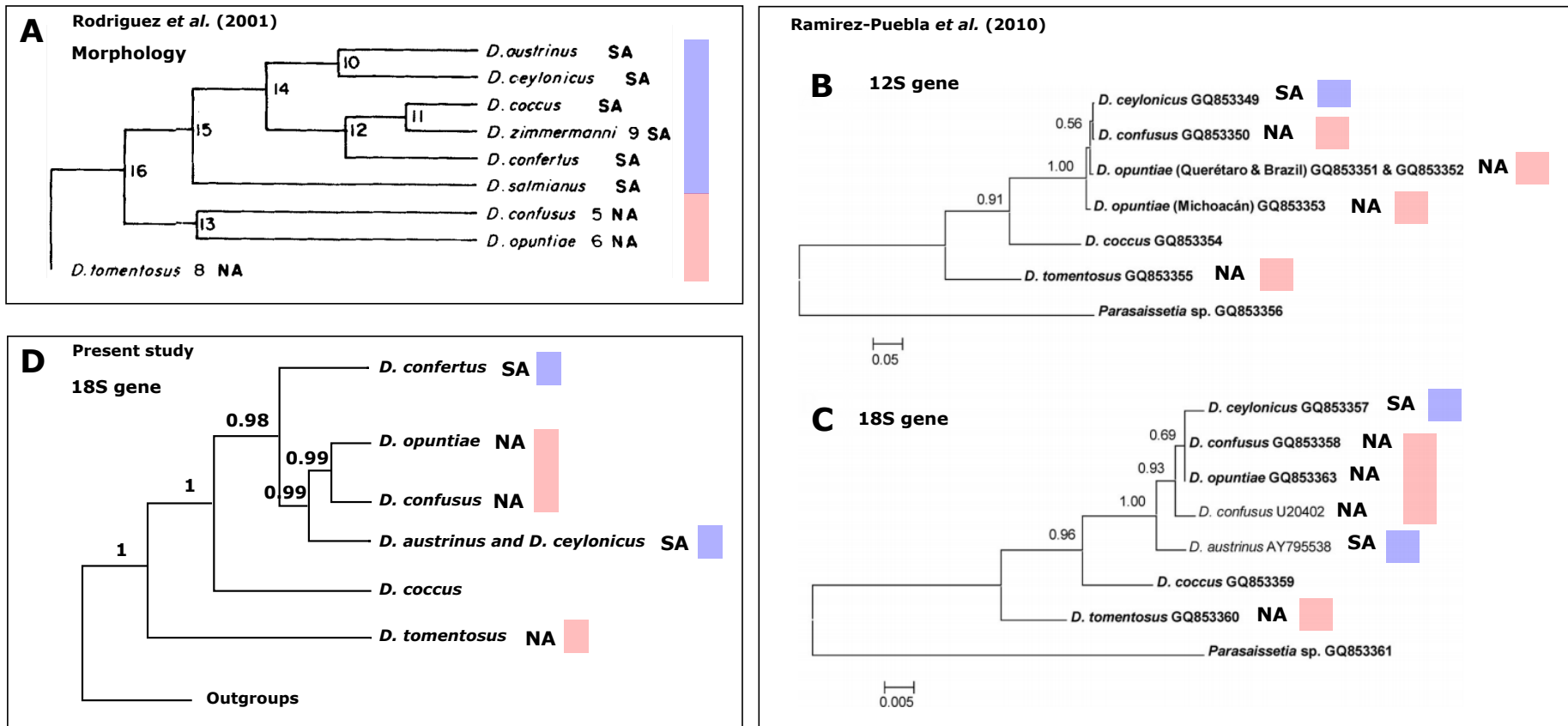
different from the other species. This is expected, because *D. tomentosus* has a considerably different life history and morphological characteristics (Pérez-Guerra and Kosztarab, 1992; Mathenge *et al.*, 2009a). The present results also agree with Rodríguez *et al.* (2001)'s finding that *D. austrinus* and *D. ceylonicus* are very closely related. These are both South American species.

Conflicting results between morphological phylogenies have been reported for this genus (Portillo and Viguera, 2006), and numerous other studies have reported on what is termed the 'morphological-molecular conflict' (Cohen, 2018). Morphological phylogenies can be misleading due to the convergent evolution of traits, as homologous morphologies are not necessarily the result of homologous genes (Cartmill, 1994). Cohen (2018), for example, states that '*morphological characters are only indirectly genealogical...and are intrinsically not capable of reporting evolutionary history*'.

The absence of a clear north-south geographic divide between the Dactylopiidae species is not unexpected, since the Cactaceae evolved in South America about 65 mya (some studies have suggested ranges between 100 (Mauseth, 1990) and 30 mya (Hershkovitz and Zimmer, 1997)) following the Gondwana split, and only later dispersed northwards (Anderson, 2001). The North and South American continents were only connected by land (the Isthmus of Panama) about 5.7 mya, and so dispersal to the north prior to that is debatable (Anderson, 2001; Portillo, 2007). It is possible that some cacti dispersed via island bridges formed by the Central American Volcanic Arc, and rafting prior to the land connection (Anderson, 2001), and it has also recently been proposed that the Isthmus of Panama may have in fact formed 13 to 10 mya (Montes *et al.*, 2015). Thus the palaeontological event known as the 'Great American Biotic Interchange' (GABI) may have occurred much earlier; although these are only hypotheses.

Based on the particular gene regions used in this study, which are also supported by the findings of Ramírez-Puebla *et al.* (2010), it may be that there were multiple dispersal events of the Dactylopiidae from South to North America at different times. And that these founder populations proceeded to speciate in North America. This raises interesting questions around the coevolution between the Cactaceae and the Dactylopiidae, and the role that human-mediated dispersal had to play in the insects' current distribution. The evolutionary history of the Cactaceae and the Dactylopiidae is still very poorly understood, and matters are made still more confusing by records of South American *D. ceylonicus* in Mexico, and North American *D. confusus* in Peru (Portillo, 2007). Disentangling human-mediated from natural dispersal events is very challenging. It would be valuable to include *D. gracilipilus*, *D. salmianus*, and *D. zimmermannii* in the current analysis

to create a complete phylogeny (at least for the currently described species), and to compare that to the phylogeny of their associated host plants. This could shed further light on host-plant evolutionary dynamics.



**Figure 4.1:** A comparison of the phylogenies produced by A) Rodríguez *et al.* (2001) based on morphological characteristics, Ramírez-Puebla *et al.* (2010) based on the B) 12S and C) 18S rRNA gene regions, and D) the present study, represented as a cladogram for the 18S region (not to scale). The values on the branches of diagrams B, C, and D are Bayesian posterior probability values. NA = North America; SA = South America.

## 4.6 Utility of DNA and molecular barcoding

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 heavily invaded by invasive cacti. This includes cases such as:

1. Identifying potential cryptic species, new lineages, and new species that could be used for the control of invasive Cactaceae in the future.
2. Identifying hybrids resulting from lineage crosses.
3. Ensuring that the most effective lineages are present at field sites to enable the highest level of control possible.
4. Distinguishing pest populations from biocontrol agent populations.

### 4.6.1 New species, cryptic species, and lineages

Countries such as South Africa and Australia are eager to source and test new agents for controlling target Cactaceae. Both Paterson *et al.* (2011) and Jones *et al.* (2016a) point out the potential need for multiple *Dactylopius* lineages to control a diverse range of species. The Dactylopiidae are notoriously difficult to identify, and, as this project has highlighted, 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. South Africa is in the process of importing what has been identified as *D. ceylonicus* for the control of *O. elata*, and has recently submitted an application of the release of *D. tomentosus* 'californica var. parkeri' for the control of *C. pallida* (Iain Paterson, pers. comm.). Up to now, the Australians have imported a total of 20 different *D. tomentosus* lineages into quarantine for the potential control of a variety of *Cylindropuntia* species (Jones *et al.*, 2018). For new agents that have not yet been released, barcoding techniques can be used to confirm their identity prior to release; as they might be completely new species! Having a genetic database to compare new sequences to, as created in this project via the R-based program 'Dacty-ID', will make this process quicker and easier. Although DNA barcoding alone is not sufficient to describe a new species, it can certainly be used to earmark potential samples for further investigation.

## 4.6.2 Identifying hybrids

Intra-specific lineages within the Dactylopiidae are known to hybridise, as discussed in Section 1.4.3. This directly affects the host specificity of hybrid populations, and thus their virulence. This 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. For example, Hoffmann *et al.* (2002) found that *D. opuntiae* ‘ficus’ and ‘stricta’ lineage crosses yielded F1 hybrids that could survive equally well on either *O. ficus-indica* or *O. stricta* (‘generalists’, as opposed to their true-bred parent ‘specialists’ that could only feed on the one preferred host species), but F2 crosses yielded offspring that contained varying combinations of both specialists and generalists. This could lead to some offspring being unable to survive if they are incompatible with their natal host plant. Alternatively, some lineage crosses might result in more effective hybrid offspring. Mathenge *et al.* (2010a), for example, found that *D. tomentosus* ‘cholla’ and ‘imbricata’ lineage crosses yielded hybrid offspring that displayed a greater fitness level on their parents’ alternative host plant than pure-bred lineages did on the same alternative host 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.

## 4.6.3 *Dactylopius opuntiae*: pest or biocontrol agent?

The *Dactylopius opuntiae* ‘ficus’ lineage (the ‘false carmine cochineal’) has become highly problematic to farmers cultivating *Opuntia ficus-indica* in areas such as Israel, Brazil, Mexico, and Morocco (Spodek *et al.*, 2014; Cruz-Rodríguez *et al.*, 2016; Torres and Giorgi, 2018). Some of these areas have reported large-scale financial losses due to the feeding damage caused by this insect. North-eastern Brazil, for example, experienced losses of up to one fifth of cultivated areas valued at some US\$100 million annually (Torres and Giorgi, 2018; Mazzeo *et al.*, 2019). Interestingly, one of the proposed hypotheses regarding the introduction and spread of this species in Brazil might also be due to a misidentification. It is postulated that populations of *D. opuntiae* were mistakenly imported instead of *D. coccus* for the creation of a red dye industry (Torres and Giorgi, 2018). Having genetic tools to distinguish between what is a pest, and what is a beneficial biocontrol agent (such as the ‘stricta’ lineage, that is not able to survive on *O. ficus-indica* (Githure *et al.*, 1999)) can be very useful. Especially since the safety reputation of biological control is at stake. For example, the *D. opuntiae*

‘stricta’ lineage has been released in Saudi Arabia to control *O. stricta*. A pest population of *D. opuntiae* ‘ficus’ is a phytosanitary pest in Israel, and was not introduced as a biocontrol agent. It is important that this pest lineage is not thought to be the same agent released in Saudi Arabia. The ‘stricta’ lineage released in Saudi Arabia is host-specific, and will not feed on cactus crop species.

## 4.7 Conclusion

This thesis aimed to create multi-locus gene phylogenies for the Dactylopiidae, and to use these data to allow for the quick identification of species and intraspecific lineages. This was achieved through the use of three DNA gene regions and two ISSRs to create a reference database. Before this study, it was not possible to tell the *D. opuntiae* ‘ficus’ and ‘stricta’ lineages apart because they are so genetically similar (and morphologically identical). The genetic identification tools arising from this work are important contributions to the biological control of invasive Cactaceae, as the correct identification of control agents is vital to its success. Even different lineages within the same species, and hybrids thereof, can result in different levels of damage to each target weed.

The taxonomic history of the Dactylopiidae is riddled with misidentifications, and so a genetic approach is a much-needed addition to the identification toolkit. This can assist in detecting new species, cryptic species, and lineages. It can also be applied to the identification of hybrids arising from lineage crosses. 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 this streamlined and accurate identification process.

# References

- Abbot, P., 2001. Individual and population variation in invertebrates revealed by Inter-simple Sequence Repeats (ISSRs). *Journal of Insect Science*, 1:8.
- Adhikari, S., Saha, S., Bandyopadhyay, T. K., and Ghosh, P., 2015. Efficiency of ISSR marker for characterization of *Cymbopogon* germplasms and their suitability in molecular barcoding. *Plant Systematics and Evolution*, 301(1):439–450.
- Ahmed, N. S. and Jaffar Ali, H. A., 2015. Numts: an impediment to DNA barcoding of Polyclinids, Tunicata. *Mitochondrial DNA*, 27(5):1–4. doi: 10.3109/19401736.2015.1018238.
- Alcaide, M., Rico, C., Ruiz, S., Soriguer, R., Muñoz, J., and Figuerola, J., 2009. Disentangling Vector-Borne Transmission Networks: A Universal DNA Barcoding Method to Identify Vertebrate Hosts from Arthropod Bloodmeals. *Public Library of Science One*, 4(9):e7092. doi: 10.1371/journal.pone.0007092.
- Alfaro, M. E., Zoller, S., and Lutzoni, F., 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Molecular Biology and Evolution*, 20(2):255–266.
- Allaby, M., 1992. *Concise Oxford Dictionary of Zoology*. Oxford University Press.
- Alston, K. P. and Richardson, D. M., 2006. The roles of habitat features, disturbance, and distance from putative source populations in structuring alien plant invasions at the urban/wildland interface on the Cape Peninsula, South Africa. *Biological Conservation*, 132(2):183–198.
- Andersen, K., Bird, K. L., Rasmussen, M., Haile, J., Breuning-Madsen, H., Kjaer, K. H., Orlando, L., Gilbert, M. T. P., and Willerslev, E., 2012. Meta-Barcoding of ‘dirt’ DNA from Soil Reflects Vertebrate Biodiversity. *Molecular Ecology*, 21(8):1966–1979. doi: 10.1111/j.1365-294X.2011.05261.x.
- Anderson, E. F., 2001. *The Cactus Family*. Timber Press, Portland, Oregon. ISBN 978-0-88192-498-5.
- Anderson, M. J., 2014. Permutational multivariate analysis of variance (PERMANOVA). *Wiley StatsRef: Statistics Reference Online*, pages 1–15.
- Anderson, N. L., Parish, N. M., Richardson, J. P., and Pearson, T. W., 1985. Comparison of African trypanosomes of different antigenic phenotypes, subspecies and life cycle stages by two-dimensional gel electrophoresis. *Molecular and Biochemical Parasitology*, 16(3):299–314.
- Anderson, N. O. and Olsen, R. T., 2015. A Vast Array of Beauty: The Accomplishments of the Father of American Ornamental Breeding, Luther Burbank. *HortScience*, 50(2):161–188.
- Anneck, D. P. and Moran, V. C., 1978. Critical reviews of biological pest control in South Africa. 2. The prickly pear, *Opuntia ficus-indica* (L.) Miller. *Journal of the Entomological Society of southern Africa*, 41(2):161–188.
- AppliedBiosystems. *DNA Fragment Analysis by Capillary Electrophoresis*. Applied Biosystems, 2014.

- Arakaki, M., Christin, P.-A., Nyffeler, R., Lendel, A., Eggli, U., Ogburn, R. M., Spriggs, E., Moore, M. J., and Edwards, E. J., 2011. Contemporaneous and Recent Radiations of the World's Major Succulent Plant Lineages. *Proceedings of the National Academy of Sciences*, 108(20):8379–8384. doi: 10.1073/pnas.1100628108.
- Archak, S., Gaikwad, A., Gautam, D., Rao, E. V., Swamy, K. R., and Karihaloo, J., 2003. Comparative assessment of DNA fingerprinting techniques (RAPD, ISSR and AFLP) for genetic analysis of cashew (*Anacardium occidentale* L.) accessions of India. *Genome*, 46(3):362–369.
- Ardura, A., Linde, A. R., Moreira, J. C., and Garcia-Vazquez, E., 2010. DNA barcoding for conservation and management of Amazonian commercial fish. *Biological Conservation*, 143(6):1438–1443.
- Arnot, D. E., Roper, C., and Bayoumi, R. A., 1993. Digital Codes from Hypervariable Tandemly Repeated DNA Sequences in the *Plasmodium falciparum* Circumsporozoite Gene Can Genetically Barcode Isolates. *Molecular and Biochemical Parasitology*, 61(1):15–24.
- Arrigo, N., Holderegger, R., and 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*, pages 155–175. Springer.
- Assefa, Y., Mitchell, A., Conlong, D. E., and Moyal, P., 2007. DNA Identification of *Busseola* (Lepidoptera: Noctuidae) Larvae in Ethiopian Sugarcane. *African Entomology*, 15(2):375–379. doi: 10.4001/1021-3589-15.2.375.
- Baker, C. S., Steel, D., Choi, Y., Lee, H., Kim, K. S., Choi, S. K., Ma, Y.-U., Hambleton, C., Psihoyos, L., Brownell, R. L., and Funahashi, N., 2010. Genetic evidence of illegal trade in protected whales links Japan with the US and South Korea. *Biology Letters*, page rsbl20100239. doi: 10.1098/rsbl.2010.0239.
- Balciunas, J. K., Burrows, D. W., and Purcell, M. F., 1996. Comparison of the physiological and realized host-ranges of a biological control agent from Australia for the control of the aquatic weed, *Hydrilla verticillata*. *Biological Control*, 7(2):148–158.
- Ball, S. L. and Armstrong, K. F., 2006. DNA Barcodes for Insect Pest Identification: A Test Case with Tussock Moths (Lepidoptera: Lymantriidae). *Canadian Journal of Forest Research*, 36(2):337–350. doi: 10.1139/x05-276.
- Ballard, J. W. O. and Whitlock, M. C., 2004. The Incomplete Natural History of Mitochondria. *Molecular Ecology*, 13(4):729–744. doi: 10.1046/j.1365-294X.2003.02063.x.
- Barker, N. P., Paterson, I., and Howis, S., 2015. “Barcoding” and ISSR data illuminate a problematic infraspecific taxonomy for *Chrysanthemoides monilifera* (Calenduleae; Asteraceae): Lessons for biocontrol of a noxious weed. *Biochemical Systematics and Ecology*, 61:541–551.
- Barr, C. M., Neiman, M., and Taylor, D. R., 2005. Inheritance and Recombination of Mitochondrial Genomes in Plants, Fungi and Animals. *The New Phytologist*, 168(1):39–50. doi: 10.1111/j.1469-8137.2005.01492.x.
- Begg, G. S., Cook, S. M., Dye, R., Ferrante, M., Franck, P., Lavigne, C., Lövei, G. L., Mansion-Vaquie, A., Pell, J. K., Petit, S., *et al.*, 2017. A functional overview of conservation biological control. *Crop Protection*, 97:145–158.
- Behera, T. K., Gaikward, A. B., Singh, A. K., and Staub, J. E., 2008. Relative efficiency of DNA markers (RAPD, ISSR and AFLP) in detecting genetic diversity of bitter melon (*Momordica charantia* L.). *Journal of the Science of Food and Agriculture*, 88(4):733–737.

- Beinart, W. and Wotshela, L., 2011. *Prickly Pear: a social history of a plant in the Eastern Cape*. Wits University Press, Johannesburg. ISBN 9781868145300.
- Ben-Dov, Y. and Hodgson, C. J., 1997. Collecting and mounting. In *World Crop Pests*, volume 7, pages 389–395. Elsevier Science. ISBN 978-0-444-89303-1.
- Ben-Dov, Y. and Marotta, S., 2001. Taxonomy and family placement of *Coccus bassi* Targioni Tozzetti, 1867 (Hemiptera: Coccoidea). *Phytoparasitica*, 29(2):169–170.
- Bensasson, D., Zhang, D. X., Hartl, D. L., and Hewitt, G. M., 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology & Evolution*, 16(6):314–321.
- Bensasson, D., Zhang, D.-X., and Hewitt, G. M., 2000. Frequent Assimilation of Mitochondrial DNA by Grasshopper Nuclear Genomes. *Molecular Biology and Evolution*, 17(3):406–415. doi: 10.1093/oxfordjournals.molbev.a026320.
- Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K., Meier, R., Winker, K., Ingram, K. K., and Das, I., 2007. Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, 22(3):148–155.
- Birch, J. L., Walsh, N. G., Cantrill, D. J., Holmes, G. D., and Murphy, D. J., 2017. Testing efficacy of distance and tree-based methods for DNA barcoding of grasses (Poaceae tribe Poeae) in Australia. *Public Library of Science One*, 12(10):e0186259. doi: 10.1371/journal.pone.0186259.
- Blaxter, M. L., 2004. The promise of a DNA taxonomy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 359(1444):669–679. doi: 10.1098/rstb.2003.1447.
- Blossey, B. and Notzold, R., 1995. Evolution of increased competitive ability in invasive nonindigenous plants: a hypothesis. *Journal of Ecology*, 83(5):887–889.
- BOLD. BarCode of Life (BOLD), 2019. URL <https://www.ncbi.nlm.nih.gov/genbank/barcode/>. Last accessed 21 November 2019.
- Bongiorni, S., Fiorenzo, P., Pippoletti, D., and Prantera, G., 2004. Inverted Meiosis and Meiotic Drive in Mealybugs. *Chromosoma*, 112(7):331–341. doi: 10.1007/s00412-004-0278-4.
- Bongiorni, S., Pasqualini, B., Taranta, M., Singh, P. B., and Prantera, G., 2007. Epigenetic Regulation of Facultative Heterochromatinisation in *Planococcus citri* via the Me(3)K9H3-HP1-Me(3)K20H4 Pathway. *Journal of Cell Science*, 120(6):1072–1080. doi: 10.1242/jcs.03412.
- Bonin, A., Bellemain, E., Bronken Eidesen, P., Pompanon, F., Brochmann, C., and Taberlet, P., 2004. How to track and assess genotyping errors in population genetics studies. *Molecular Ecology*, 13(11):3261–3273.
- Bonin, A., Ehrich, D., and Manel, S., 2007. Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. *Molecular Ecology*, 16(18):3737–3758.
- Bornet, B. and Branchard, M., 2001. Nonanchored inter simple sequence repeat (ISSR) markers: reproducible and specific tools for genome fingerprinting. *Plant Molecular Biology Reporter*, 19(3):209–215.
- Bornet, B., Antoine, E., Bardouil, M., and Marcaillou-Le Baut, C., 2004. ISSR as new markers for genetic characterization and evaluation of relationships among phytoplankton. *Journal of Applied Phycology*, 16(4):285–290.

- Briski, E., Cristescu, M. E., Bailey, S. A., and MacIsaac, H. J., 2011. Use of DNA Barcoding to Detect Invertebrate Invasive Species from Diapausing Eggs. *Biological Invasions*, 13(6):1325–1340. doi: 10.1007/s10530-010-9892-7.
- Brooks, M. L., D'Antonio, C. M., Richardson, D. M., Grace, J. B., Keeley, J. E., DiTomaso, J. M., Hobbs, R. J., Pellant, M., and Pyke, D., 2004. Effects of Invasive Alien Plants on Fire Regimes. *BioScience*, 54(7):677–688. doi: 10.1641/0006-3568(2004)054[0677:EOLAPO]2.0.CO;2.
- Brown, C. J., Macdonald, I. A. W., and 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., and Cruickshank Robert H., 2012. Spider: An R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. *Molecular Ecology Resources*, 12(3):562–565. doi: 10.1111/j.1755-0998.2011.03108.x.
- Brown, S. and Nur, U., 1964. Heterochromatic Chromosomes in the Coccids. *Science*, 145(3628):130–136. doi: 10.1126/science.145.3628.130.
- Bruford, M. W., Hanotte, O., Brookfield, J. F. Y., and Burke, T., 1992. Single locus and multilocus DNA fingerprinting. In Hoazel, C. A. R., editor, *Molecular Genetics Analysis of Populations: A Practical Approach*, pages 225–269. Oxford University Press, New York.
- Brun, L. O., Borsa, P., Gaudichon, V., Stuart, J. J., Aronstein, K., Coustau, C., and French-Constant, R. H., 1995. 'Functional' haplodiploidy. *Nature*, 374(6522):506. doi: 10.1038/374506a0.
- Brutsch, M. O. and Zimmermann, H. G., 1993. The Prickly Pear (*Opuntia Ficus-indica* [Cactaceae]) in South Africa: Utilization of the Naturalized Weed, and of the Cultivated Plants. *Economic Botany*, 47(2):154–162. doi: 10.1007/BF02862018.
- Buccellato, L., Byrne, M. J., Fisher, J. T., and Witkowski, E. T., 2019. Post-release evaluation of a combination of biocontrol agents on crofton weed: testing extrapolation of greenhouse results to field conditions. *BioControl*, 64(4):457–468.
- Buchner, P., 1965. *Endosymbiosis of animals with plant microorganisms*. Interscience publishers, New York.
- Buglia, G. L. and Ferraro, M., 2004. Germline cyst development and imprinting in male mealybug *Planococcus citri*. *Chromosoma*, 113(6):284–294. doi: 10.1007/s00412-004-0317-1.
- Buhay, J. E., 2009. COI-like Sequences Are Becoming Problematic in Molecular Systematic and DNA Barcoding Studies. *Journal of Crustacean Biology*, 29(1):96–110. doi: 10.1651/08-3020.1.
- Bull, J. J., 1979. An Advantage for the Evolution of Male Haploidy and Systems with Similar Genetic Transmission. *Heredity*, 43(3):361–381. doi: 10.1038/hdy.1979.88.
- Burd, J. D. and Porter, D. R., 2006. Biotypic diversity in greenbug (Hemiptera: Aphididae): characterizing new virulence and host associations. *Journal of Economic Entomology*, 99(3):959–965.
- Burt, A. and Trivers, R., 2009. *Genes in Conflict: The Biology of Selfish Genetic Elements*. Harvard University Press. ISBN 978-0-674-02911-8.
- Byrne, M., Machimane, Z., Paterson, I., Witt, A., Cowie, B., and Venter, N., 2018. Cochineal and cactus: are new associations biocontrol winners? In *XV International Symposium on Biological Control of Weeds*, Hotel Bellevue, Engelberg, Switzerland. ISBCW.

- Cameron, S. L., 2014. Insect Mitochondrial Genomics: Implications for Evolution and Phylogeny. *Annual Review of Entomology*, 59(1):95–117. doi: 10.1146/annurev-ento-011613-162007.
- Campana, M. G., Robles García, N. M., and Tuross, N., 2015. America's Red Gold: Multiple Lineages of Cultivated Cochineal in Mexico. *Ecology and Evolution*, 5(3):607–617. doi: 10.1002/ece3.1398.
- Camus, M. F., Clancy, D. J., and Dowling, D. K., 2012. Mitochondria, Maternal Inheritance, and Male Aging. *Current Biology*, 22(18):1717–1721. doi: 10.1016/j.cub.2012.07.018.
- Cartmill, M., 1994. A critique of homology as a morphological concept. *American Journal of Physical Anthropology*, 94(1):115–123.
- Casu, M., Maltagliati, F., Cossu, P., Lai, T., Galletti, M. C., Castelli, A., and Commito, J. A., 2005. Fine-grained spatial genetic structure in the bivalve *Gemma gemma* from Maine and Virginia (USA), as revealed by Inter-Simple Sequence Repeat markers. *Journal of Experimental Marine Biology and Ecology*, 325(1):46–54.
- Casu, M., Casu, D., Lai, T., Cossu, P., and Curini-Galletti, M., 2006. Inter-simple sequence repeat markers reveal strong genetic differentiation among populations of the endangered mollusc *Patella ferruginea* (Gastropoda: Patellidae) from two Sardinian marine protected areas. *Marine Biology*, 149(5):1163–1174.
- Cavalier-Smith, T., 1978. Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate, and the solution of the DNA C-value paradox. *Journal of Cell Science*, 34:247–278.
- Cawthorn, D.-M., Steinman, H. A., and Witthuhn, R. C., 2012. Evaluation of the 16S and 12S rRNA genes as universal markers for the identification of commercial fish species in South Africa. *Gene*, 491(1):40–48.
- CBOL, 2009. A DNA barcode for land plants. In *Proceedings of the National Academy of Sciences*, pages 12794–12797, 2009. doi: 10.1073/pnas.0905845106.
- Chandra, H. S., 1962. Inverse Meiosis in Triploid Females of the Mealy Bug, *Planococcus citri*. *Genetics*, 47(10):1441–1454.
- Chatterjee, S. and Mohandas, T., 2003. Identification of ISSR markers associated with productivity traits in silkworm, *Bombyx mori* L. *Genome*, 46(3):438–447.
- Chávez-Moreno, C. K., Tecante, A., and Casas, A., 2009. The *Opuntia* (Cactaceae) and *Dactylopius* (Hemiptera: Dactylopiidae) in Mexico: A Historical Perspective of Use, Interaction and Distribution. *Biodiversity and Conservation*, 18(13):3337–3355. doi: 10.1007/s10531-009-9647-x.
- Chávez-Moreno, C., Tecante, A., Casas, A., and Claps, L., 2011. Distribution and Habitat in Mexico of *Dactylopius* Costa (Hemiptera: Dactylopiidae) and their Cacti Hosts (Cactaceae: Opuntioideae). *Neotropical Entomology*, 40(1):62–71. doi: 10.1590/S1519-566X2011000100009.
- Christenhusz, M. J. and Byng, J. W., 2016. The number of known plants species in the world and its annual increase. *Phytotaxa*, 261(3):201–217.
- Claps, L. and de Haro, M., 2001. Coccoidea (Insecta: Hemiptera) Associated With Cactaceae in Argentina. *Journal of the Professional Association for the Cactus Development*, 4:77–83.
- Claridge, M. and Den Hollander, J., 1983. The biotype concept and its application to insect pests of agriculture. *Crop Protection*, 2(1):85–95. doi: 10.1016/0261-2194(83)90028-5.

- Clarke, K. R., 1993. Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology*, 18(1):117–143.
- Clayton, D. A., 2003. Mitochondrial DNA Replication: What We Know. *IUBMB Life*, 55(4-5):213–217. doi: 10.1080/1521654031000134824.
- Clement, M., Posada, D., and Crandall, K. A., 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, 9(10):1657–1659.
- Cohen, B. L., 2018. Match and mismatch of morphological and molecular phylogenies: causes, implications, and new light on cladistics. *Zoological Journal of the Linnean Society*, 184(2):516–527.
- Collier, T. and Van Steenwyk, R., 2004. A critical evaluation of augmentative biological control. *Biological Control*, 31(2):245–256.
- Cong, Q., Shen, J., Borek, D., Robbins, R. K., Opler, P. A., Otwinowski, Z., and Grishin, N. V., 2017. When COI barcodes deceive: complete genomes reveal introgression in hairstreaks. *Proceedings. Biological Sciences*, 284(1848):20161735. doi: 10.1098/rspb.2016.1735.
- Cook, L. G., 2000. Extraordinary and Extensive Karyotypic Variation: A 48-Fold Range in Chromosome Number in the Gall-Inducing Scale Insect *Apiomorpha* (Hemiptera: Coccoidea: Eriococcidae). *Genome*, 43(2):255–263.
- Cordero, R. D., Sánchez-Ramírez, S., and Currie, D. C., 2017. DNA barcoding of aquatic insects reveals unforeseen diversity and recurrent population divergence patterns through broad-scale sampling in northern Canada. *Polar Biology*, 40(8):1687–1695. doi: 10.1007/s00300-016-2062-3.
- Costa, R., Pereira, G., Garrido, I., Tavares-de Sousa, M. M., and Espinosa, F., 2016. Comparison of RAPD, ISSR, and AFLP molecular markers to reveal and classify orchardgrass (*Dactylis glomerata* L.) germplasm variations. *Public Library of Science One*, 11(4):e0152972.
- Cox, A. J. and Hebert, P. D., 2001. Colonization, Extinction, and Phylogeographic Patterning in a Freshwater Crustacean. *Molecular Ecology*, 10(2):371–386.
- Cristofaro, M., Sale, F., Campobasso, G., Knutson, L., and Sbordoni, V., 1998. Biology and host preference of *Nephoteryx divisella* (Lepidoptera: Pyralidae): Candidate agent for biological control of leafy spurge complex in North America. *Environmental Entomology*, 27(3):731–735.
- Crooks, J. A., 2002. Characterizing Ecosystem-Level Consequences of Biological Invasions: The Role of Ecosystem Engineers. *Oikos*, 97(2):153–166. doi: 10.1034/j.1600-0706.2002.970201.x.
- Cruz-Rodríguez, J., González-Machorro, E., Villegas González, A., Rodríguez Ramírez, M., and Mejía Lara, F., 2016. Autonomous biological control of *Dactylopius opuntiae* (Hemiptera: Dactylopiidae) in a prickly pear plantation with ecological management. *Environmental Entomology*, 45(3):642–648.
- Ćukušić, A., Ćuk, R., Previšić, A., Podnar, M., Delić, A., and Kučinić, M., 2017. DNA barcoding and first records of two rare *Adicella* species (Trichoptera: Leptoceridae) in Croatia. *Biologia*, 72(7). doi: 10.1515/biolog-2017-0087.
- Cullen, J., 1995. Predicting effectiveness: fact and fantasy. In Delfosse, E. and Scott, R., editors, *Proceedings of the VIII International Symposium on Biological Control of Weeds*, pages 103–109. Agriculture Canada, 1995.

- Culliney, T. W., 2005. Benefits of classical biological control for managing invasive plants. *Critical Reviews in Plant Sciences*, 24(2):131–150.
- Dallai, R., Fanciulli, P. P., and Frati, F., 2000. Aberrant Spermatogenesis and the Peculiar Mechanism of Sex Determination in Symphypleonan Collembola (Insecta). *The Journal of Heredity*, 91(5):351–358.
- D'Antonio, C. M. and Vitousek, P. M., 1992. Biological Invasions by Exotic Grasses, the Grass/-Fire Cycle, and Global Change. *Annual Review of Ecology and Systematics*, 23(1):63–87. doi: 10.1146/annurev.es.23.110192.000431.
- Darlington, C. D. and Mather, K., 1952. *The Elements of Genetics*. The Macmillan Company.
- Darriba, D., Taboada, G. L., Doallo, R., and Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, 9(8):772–772. doi: 10.1038/nmeth.2109.
- Day, M., Jones, P., and Holtkamp, R., 2015. Application for the release of the cochineal *Dactylopius tomentosus* ('fulgida' biotype) for the biological control of *Cylindropuntia fulgida* var. *mamillata* (Cactaceae). Technical report, Biosecurity Queensland Department of Agriculture, Fisheries and Forestry.
- de Beer, H., 1986. *Rosea* cactus. Technical report, Department of Agriculture and Water Supply, Pretoria.
- de la Filia, A. G., Bain, S. A., and Ross, L., 2015. Haplodiploidy and the Reproductive Ecology of Arthropods. *Current Opinion in Insect Science*, 9:36–43. doi: 10.1016/j.cois.2015.04.018.
- de Lange, W. J. and van Wilgen, B. W., 2010. An Economic Assessment of the Contribution of Biological Control to the Management of Invasive Alien Plants and to the Protection of Ecosystem Services in South Africa. *Biological Invasions*, 12(12):4113–4124. doi: 10.1007/s10530-010-9811-y.
- De Lotto, G., 1974. On the status and identity of the cochineal insects (Homoptera: Coccoidea: Dactylopiidae). *Journal of the Entomological Society of Southern Africa*, 37(1):167–193.
- De Mandal, S., Chhakchhuak, L., Gurusubramanian, G., and Kumar, N. S., 2014. Mitochondrial markers for identification and phylogenetic studies in insects - A Review. *DNA Barcodes*, 2(1). doi: 10.2478/dna-2014-0001.
- de Vienne, D. M., Giraud, T., and Martin, O. C., 2007. A congruence index for testing topological similarity between trees. *Bioinformatics*, 23(23):3119–3124.
- Deagle, B. E., Jarman, S. N., Pemberton, D., and Gales, N. J., 2005. Genetic Screening for Prey in the Gut Contents from a Giant Squid (*Architeuthis* sp.). *The Journal of Heredity*, 96(4):417–423. doi: 10.1093/jhered/esi036.
- Dean, W. R. J. and Milton, S. J., 2000. Directed Dispersal of *Opuntia* Species in the Karoo, South Africa: Are Crows the Responsible Agents? *Journal of Arid Environments*, 45(4):305–314. doi: 10.1006/jare.2000.0652.
- Department of Environmental Affairs, 2016. National environmental management: Biodiversity act, 2004 (act no. 10 of 2004) alien and invasive species lists, 2016. Technical report, DEA.
- D'Errico, I., Gadaleta, G., and Saccone, C., 2004. Pseudogenes in metazoa: origin and features. *Briefings in Functional Genomics & Proteomics*, 3(2):157–67.

- Dhileepan, K., Senaratne, K. W., and Raghu, S., 2006. A systematic approach to biological control agent exploration and prioritisation for prickly acacia (*Acacia nilotica* ssp. *indica*). *Australian Journal of Entomology*, 45(4):303–307.
- Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J.-M., Taly, J.-F., and Notredame, C., 2011. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Research*, 39(suppl\_2): W13–W17.
- Diehl, S. R. and Bush, G. L., 1984. An Evolutionary and Applied Perspective of Insect Biotypes. *Annual Review of Entomology*, 29(1):471–504. doi: 10.1146/annurev.en.29.010184.002351.
- Dodd, A., 1940. The biological control of prickly pear in Australia. *Commonwealth Bureau of Pastures and Field Crops*, 27:131–143.
- Domingo-Roura, X., Marmi, J., Ferrando, A., López-Giráldez, F., Macdonald, D. W., and Jansman, H. A. H., 2006. Badger Hair in Shaving Brushes Comes from Protected Eurasian Badgers. *Biological Conservation*, 128(3):425–430. doi: 10.1016/j.biocon.2005.08.013.
- Donkin, R. A., 1977. Spanish Red: An Ethnogeographical Study of Cochineal and the *Opuntia* Cactus. *Transactions of the American Philosophical Society*, 67(5):1. doi: 10.2307/1006195.
- Douady, C. J., Delsuc, F., Boucher, Y., Doolittle, W. F., and Douzery, E. J., 2003. Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. *Molecular Biology and Evolution*, 20(2):248–254.
- Douglas, A. E., 1998. Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria *Buchnera*. *Annual Review of Entomology*, 43:17–37. doi: 10.1146/annurev.ento.43.1.17.
- Douglas, A. E., 2006. Phloem-Sap Feeding by Animals: Problems and Solutions. *Journal of Experimental Botany*, 57(4):747–754. doi: 10.1093/jxb/erj067.
- Douglas, A. E., 2015. Multiorganismal insects: diversity and function of resident microorganisms. *Annual Review of Entomology*, 60:17–34. doi: 10.1146/annurev-ento-010814-020822.
- Downie, D. A., 2010. Baubles, Bangles, and Biotypes: A Critical Review of the use and Abuse of the Biotype Concept. *Journal of Insect Science*, 10(176):1–18. doi: 10.1673/031.010.14136.
- Drake, J. W., Charlesworth, B., Charlesworth, D., and Crow, J. F., 1998. Rates of spontaneous mutation. *Genetics*, 148(4):1667–1686.
- Drès, M. and Mallet, J., 2002. Host races in plant-feeding insects and their importance in sympatric speciation. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 357(1420):471–492.
- du Buy, H. G. and Riley, F. L., 1967. Hybridization between the nuclear and kinetoplast DNAs of *Leishmania enriettii* and between nuclear and mitochondrial DNAs of mouse liver. *Proceedings of the National Academy of Sciences of the United States of America*, 57(3):790–797.
- Dugard, P., Todman, J., and Staines, H., 2010. *Approaching multivariate analysis. A practical introduction.* CRC Press, Taylor and Francis, Routledge, New York.
- Dye, P. and Jarmain, C., 2004. Water use by black wattle (*Acacia mearnsii*): implications for the link between removal of invading trees and catchment streamflow response. *South African Journal of Science*, page 5.

- Eastop, V., 1973. Biotypes of aphids. *Bulletin of the Entomological Society of New Zealand*, 2:40–51.
- Edmunds, G. F. and Alstad, D. N., 1978. Coevolution in insect herbivores and conifers. *Science*, 199 (4332):941–945.
- Edwards, E. J., Nyffeler, R., and Donoghue, M. J., 2005. Basal Cactus Phylogeny: Implications of *Pereskia* (Cactaceae) Paraphyly for the Transition to the Cactus Life Form. *American Journal of Botany*, 92(7):1177–1188. doi: 10.3732/ajb.92.7.1177.
- Edwards, S. V., 2009. Is a new and general theory of molecular systematics emerging? *Evolution*, 63 (1):1–19. doi: 10.1111/j.1558-5646.2008.00549.x.
- Eggl, U., 1993. *Glossary of botanical terms with special reference to succulent plants, including German equivalents*. British Cactus & Succulent Society.
- Eisner, T., Nowicki, S., Goetz, M., and Meinwald, J., 1980. Red Cochineal Dye (Carminic Acid): Its Role in Nature. *Science, New Series*, 208(4447):1039–1042.
- Ermakov, O. A., Simonov, E., Surin, V. L., Titov, S. V., Brandler, O. V., Ivanova, N. V., and Borisenko, A. V., 2015. Implications of Hybridization, NUMTs, and Overlooked Diversity for DNA Barcoding of Eurasian Ground Squirrels. *Public Library of Science One*, 10(1):e0117201. doi: 10.1371/journal.pone.0117201.
- Evanno, G., Regnaut, S., and Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14(8):2611–2620.
- Evans, H. C., 1997. *Parthenium hysterophorus*: A Review of Its Weed Status and the Possibilities for Biological Control. *Biocontrol News and Information*, 18(3):89–98.
- Evans, K. M., Wortley, A. H., and Mann, D. G., 2007. An assessment of potential diatom barcode genes (cox1, rbcL, 18S and ITS rDNA) and their effectiveness in determining relationships in *Sellaphora* (Bacillariophyta). *Protist*, 158(3):349–364.
- Falush, D., Stephens, M., and Pritchard, J. K., 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, 164(4):1567–1587.
- Felsenstein, J., 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution*, 17(6):368–376. doi: 10.1007/BF01734359.
- Ficetola, G. F., Miaud, C., Pompanon, F., and Taberlet, P., 2008. Species Detection Using Environmental DNA from Water Samples. *Biology Letters*, 4(4):423–425. doi: 10.1098/rsbl.2008.0118.
- Fischbein, D., Lantschner, M. V., and Corley, J. C., 2019. Modelling the distribution of forest pest natural enemies across invaded areas: Towards understanding the influence of climate on parasitoid establishment success. *Biological Control*, 132:177–188.
- Floyd, R., Lima, J., deWaard, J., Humble, L., and Hanner, R., 2010. Common Goals: Policy Implications of DNA Barcoding as a Protocol for Identification of Arthropod Pests. *Biological Invasions*, 12(9): 2947–2954. doi: 10.1007/s10530-010-9709-8.
- Fontaine, K. M., Cooley, J. R., and Simon, C., 2007. Evidence for Paternal Leakage in Hybrid Periodical Cicadas (Hemiptera: *Magicalicada* Spp.). *Public Library of Science One*, 2(9):e892. doi: 10.1371/journal.pone.0000892.

- Foxcroft, L. and Hoffmann, J., 2000. Dispersal of *Dactylopius opuntiae* Cockerell (Homoptera: Dactylopiidae), a biological control agent of *Opuntia stricta* (Haworth.) Haworth.(Cactaceae) in the Kruger National Park. *Koedoe*, 43(2):1–5.
- Foxcroft, L. C., Rouget, M., Richardson, D. M., and Mac Fadyen, S., 2004. Reconstructing 50 years of *Opuntia stricta* invasion in the Kruger National Park, South Africa: environmental determinants and propagule pressure. *Diversity and Distributions*, 10(5-6):427–437.
- Frank, S. and Hurst, L., 1996. Mitochondria and Male Disease. *Nature*, 383:224. doi: 10.1038/383224a0.
- Freeman, S., 2005. *Biological Science*. Pearson, 2nd edition.
- Frézal, L. and Leblois, R., 2008. Four Years of DNA Barcoding: Current Advances and Prospects. *Infection, Genetics and Evolution*, 8(5):727–736. doi: 10.1016/j.meegid.2008.05.005.
- Fried, G., Chauvel, B., Reynaud, P., and Sache, I., 2017. Decreases in Crop Production by Non-native Weeds, Pests, and Pathogens. In Vilà, M. and Hulme, P. E., editors, *Impact of Biological Invasions on Ecosystem Services*, chapter 6, pages 83–101. Springer International Publishing, Cham. doi: 10.1007/978/3/319/45121/3/6.
- Frost, S., 1954. The numerical relationships between phytophagous insects and their hosts. *The Scientific Monthly*, 79(1):10–12.
- Funk, D. J. and Omland, K. E., 2003. Species-Level Paraphyly and Polyphyly: Frequency, Causes, and Consequences, with Insights from Animal Mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics*, 34(1):397–423. doi: 10.1146/annurev.ecolsys.34.011802.132421.
- Galbács, Z., Molnar, S., Halász, G., Kozma, P., Hoffmann, S., Kovacs, L., Veres, A., Galli, Z., Szőke, A., Heszky, L., and Kiss, E., 2009. Identification of grapevine cultivars using microsatellite-based DNA barcodes. *VITIS-Journal of Grapevine Research*, 48(1):17.
- Galitski, T., Saldanha, A. J., Styles, C. A., Lander, E. S., and Fink, G. R., 1999. Ploidy Regulation of Gene Expression. *Science*, 285(5425):251–254. doi: 10.1126/science.285.5425.251.
- Garcia, A. A., Benchimol, L. L., Barbosa, A. M., Geraldi, I. O., Souza Jr, C. L., and de Souza, A. P., 2004. Comparison of RAPD, RFLP, AFLP and SSR markers for diversity studies in tropical maize inbred lines. *Genetics and Molecular Biology*, 27(4):579–588.
- Gaskin, J. F., Bon, M.-C., Cock, M. J. W., Cristofaro, M., Biase, A. D., De Clerck-Floate, R., Ellison, C. A., Hinz, H. L., Hufbauer, R. A., Julien, M. H., and Sforza, R., 2011. Applying Molecular-Based Approaches to Classical Biological Control of Weeds. *Biological Control*, 58(1):1–21. doi: 10.1016/j.biocontrol.2011.03.015.
- Gavrilov, I. A., 2007. A catalog of chromosome numbers and genetic systems of scale insects (Homoptera: Coccinea) of the world. *Israel Journal of Entomology*, 37:1–45.
- Gellissen, G., Bradfield, J. Y., White, B. N., and Wyatt, G. R., 1983. Mitochondrial DNA sequences in the nuclear genome of a locust. *Nature*, 301(5901):631–634. doi: 10.1038/301631a0.
- Gentekaki, E. and Lynn, D., 2012. Spatial Genetic Variation, Phylogeography and Barcoding of the Peritrichous Ciliate *Carchesium polypinum*. *European Journal of Protistology*, 48(4):305–313. doi: 10.1016/j.ejop.2012.04.001.

- Gerth, M., Geisbler, A., and Bleidorn, C., 2011. *Wolbachia* infections in bees (*Anthophila*) and possible implications for DNA barcoding. *Systematics and Biodiversity*, 9(4):319–327. doi: 10.1080/14772000.2011.627953.
- Gibson, A., Gowri-Shankar, V., Higgs, P. G., and Rattray, M., 2004. A Comprehensive Analysis of Mammalian Mitochondrial Genome Base Composition and Improved Phylogenetic Methods. *Molecular Biology and Evolution*, 22(2):251–264. doi: 10.1093/molbev/msi012.
- Gilreath, M. E. and Smith, J. W., 1987. Bionomics of *Dactylopius confusus* (Homoptera: Dactylopiidae). *Annals of the Entomological Society of America*, 80(6):768–774. doi: 10.1093/aesa/80.6.768.
- Githure, C., Zimmermann, H., and Hoffmann, J., 1999. Host specificity of biotypes of *Dactylopius opuntiae* (Cockerell) (Hemiptera: Dactylopiidae): prospects for biological control of *Opuntia stricta* (Haworth) (Cactaceae) in Africa. *African Entomology*, 7(1):43–48.
- Goeden, R., 1983. Critique and revision of Harris' scoring system for selection of insect agents in biological control of weeds. *Protection Ecology*, 5(4):287–301.
- Gon, S. M. and Price, E. O., 1984. Invertebrate domestication: behavioral considerations. *Bioscience*, 34(9):575–579.
- Goolsby, J. A., De Barro, P. J., Makinson, J. R., Pemberton, R. W., Hartley, D. M., and Frohlich, D. R., 2006. Matching the origin of an invasive weed for selection of a herbivore haplotype for a biological control programme. *Molecular Ecology*, 15(1):287–297.
- Gray, M. W., Burger, G., and Lang, B. F., 1999. Mitochondrial Evolution. *Science*, 283(5407):1476–1481. doi: 10.1126/science.283.5407.1476.
- Greathead, D., 1995. Benefits and risks of classical biological control. In Hokkanen, H. and Lynch, J., editors, *Biological control: benefits and risks*, chapter 5, pages 53–63. Cambridge University Press, Cambridge, UK.
- Greenfield, A. B., 2005. *A Perfect Red: Empire, Espionage, and the Quest for the Color of Desire*. Transworld Publishers, London.
- Griffith, M. P., 2003. Using molecular evidence to elucidate reticulate evolution in *Opuntia* (Cactaceae). *Madroño*, 50(3):162–169.
- Griffith, M. P. and Porter, J. M., 2009. Phylogeny of Opuntioideae (Cactaceae). *International Journal of Plant Sciences*, 170(1):107–116.
- Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Léger, P., Lepais, O., Lepoittevin, C., Malausa, T., Revardel, E., Salin, F., and Petit, R., 2011. Current trends in microsatellite genotyping. *Molecular Ecology Resources*, 11(4):591–611.
- Guicking, D., Griffiths, R. A., Moore, R. D., Joger, U., and Wink, M., 2006. Introduced alien or persecuted native? Resolving the origin of the viperine snake (*Natrix maura*) on Mallorca. *Biodiversity & Conservation*, 15(9):3045–3054.
- Guindon, S. and Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, 52(5):696–704.
- Gullan, P. J. and Cranston, P. S., 2014. *The insects: an outline of entomology*. John Wiley & Sons, West Sussex, UK, 5th edition.

- Gullan, P. J. and Kosztarab, M., 1997. Adaptations in Scale Insects. *Annual Review of Entomology*, 42(1):23–50. doi: 10.1146/annurev.ento.42.1.23.
- Gunn, B. H., 1979. *Dispersal of the Cochineal Insect Dactylopius Austrinus de Lotto (Homoptera:Dactylopiidae)*. PhD thesis, Rhodes University.
- Gutiérrez, M. A. C., Vivero, R. J., Vélez, I. D., Porter, C. H., and Uribe, S., 2014. DNA Barcoding for the Identification of Sand Fly Species (Diptera, Psychodidae, Phlebotominae) in Colombia. *Public Library of Science One*, 9(1):e85496. doi: 10.1371/journal.pone.0085496.
- Gyllenstein, U., Wharton, D., Josefsson, A., and Wilson, A. C., 1991. Paternal inheritance of mitochondrial DNA in mice. *Nature*, 352(6332):255–257. doi: 10.1038/352255a0.
- Haag-Liautard, C., Coffey, N., Houle, D., Lynch, M., Charlesworth, B., and Keightley, P. D., 2008. Direct Estimation of the Mitochondrial DNA Mutation Rate in *Drosophila melanogaster*. *Public Library of Science Biology*, 6(8). doi: 10.1371/journal.pbio.0060204.
- Haig, D., 1993a. The evolution of unusual chromosomal systems in coccoids: extraordinary sex ratios revisited. *Journal of Evolutionary Biology*, 6(1):69–77. doi: 10.1046/j.1420-9101.1993.6010069.x.
- Haig, D., 1993b. The evolution of unusual chromosomal systems in sciarid flies: intragenomic conflict and the sex ratio. *Journal of Evolutionary Biology*, 6(2):249–261. doi: 10.1046/j.1420-9101.1993.6020249.x.
- Haig, S. M., Mace, T. R., and Mullins, T. D., 2003. Parentage and relatedness in polyandrous comb-crested jacanas using ISSRs. *Journal of Heredity*, 94(4):302–309.
- Hall, T., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In *Nucleic Acids Symposium Series*, pages 95–98, Raleigh, North Carolina, USA. Oxford University Press.
- Hammer, ., Harper, D. A. T., and Ryan, P. D., 2001. PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontologia Electronica*, 4:9.
- Hardy, C. M., Adams, M., Jerry, D. R., Morgan, M. J., Hartley, D. M., *et al.*, 2011. DNA barcoding to support conservation: species identification, genetic structure and biogeography of fishes in the MurrayDarling River Basin, Australia. *Marine and Freshwater Research*, 62(8):887–901.
- Harris, P., 1991. Classical biocontrol of weeds: its definitions, selection of effective agents, and administrative-political problems. *The Canadian Entomologist*, 123:827–849.
- Harris, P., 1973. The selection of effective agents for the biological control of weeds. *The Canadian Entomologist*, 105(12):1495–1503.
- Harrison, R. G. and Larson, E. L., 2014. Hybridization, Introgression, and the Nature of Species Boundaries. *Journal of Heredity*, 105(S1):795–809. doi: 10.1093/jhered/esu033.
- Häuser, C. L., 1987. The debate about the biological species concept—a review. *Journal of Zoological Systematics and Evolutionary Research*, 25(4):241–257.
- Hausmann, A., Godfray, H. C. J., Huemer, P., Mutanen, M., Rougerie, R., van Nieukerken, E. J., Ratnasingham, S., and Hebert, P. D. N., 2013. Genetic Patterns in European Geometrid Moths Revealed by the Barcode Index Number (BIN) System. *Public Library of Science One*, 8(12):e84518. doi: 10.1371/journal.pone.0084518.

- Havemann, N., Gossner, M. M., Hendrich, L., Morinière, J., Niedringhaus, R., Schäfer, P., and Raupach, M. J., 2018. From water striders to water bugs: the molecular diversity of aquatic Heteroptera (Gerromorpha, Nepomorpha) of Germany based on DNA barcodes. *PeerJ*, 6:e4577. doi: 10.7717/peerj.4577.
- Hawlitsek, O., Morinière, J., Lehmann, G. U. C., Lehmann, A. W., Kropf, M., Dunz, A., Glaw, F., Detcharoen, M., Schmidt, S., Hausmann, A., Szucsich, N. U., Caetano-Wyler, S. A., and Haszprunar, G., 2017. DNA barcoding of crickets, katydids and grasshoppers (Orthoptera) from Central Europe with focus on Austria, Germany and Switzerland. *Molecular Ecology Resources*, 17(5):1037–1053. doi: 10.1111/1755-0998.12638.
- Haye, T., Kuhlmann, U., Goulet, H., and Mason, P., 2006. Controlling Lygus plant bugs (Heteroptera: Miridae) with European *Peristenus relictus* (Hymenoptera: Braconidae) in Canada—risky or not? *Bulletin of Entomological Research*, 96(2):187–196.
- Hazkani-Covo, E., Zeller, R. M., and Martin, W., 2010. Molecular Poltergeists: Mitochondrial DNA Copies (numts) in Sequenced Nuclear Genomes. *Public Library of Science Genetics*, 6(2):e1000834. doi: 10.1371/journal.pgen.1000834.
- Heard, T. A., 1999. Concepts in insect host-plant selection behavior and their application to host specificity testing. In *Host Specificity of Exotic Arthropod Biological Control Agents: the Biological Basis for Improvement in Safety, Proceedings of Session: X International Symposium on Biological Control of Weeds*, pages 1–10, 1999.
- Heard, T. A., 2002. Host specificity testing of biocontrol agents of weeds. In *Proceedings of a Workshop on Biological Control of Invasive Plants in Native Hawaiian Ecosystems. Pacific Cooperative Studies Unit, University of Hawaii at Manoa. Edited by Smith, CW, Denslow, JE, and Hight, SD Technical Report*, volume 129, pages 21–29, 2002.
- Hebert, P. D. N., Cywinska, A., Ball, S. L., and deWaard, J. R., 2003a. Biological Identifications through DNA Barcodes. *Proceedings of the Royal Society B: Biological Sciences*, 270(1512):313–321. doi: 10.1098/rspb.2002.2218.
- Hebert, P. D. N., Ratnasingham, S., and de Waard, J. R., 2003b. Barcoding Animal Life: Cytochrome c Oxidase Subunit 1 Divergences among Closely Related Species. *Proceedings of the Royal Society B: Biological Sciences*, 270(Suppl.1):S96–S99. doi: 10.1098/rsbl.2003.0025.
- Henderson, L., 2001. *Alien Weeds and Invasive Plants: a complete guide to declared weeds and invaders in South Africa*. Agricultural Research Council, Pretoria.
- Hendrich, L., Morinière, J., Haszprunar, G., Hebert, P. D. N., Hausmann, A., Köhler, F., and Balke, M., 2015. A comprehensive DNA barcode database for Central European beetles with a focus on Germany: adding more than 3500 identified species to BOLD. *Molecular Ecology Resources*, 15(4): 795–818. doi: 10.1111/1755-0998.12354.
- Herrick, G. and Seger, J., 1999. Imprinting and Paternal Genome Elimination in Insects. In Ohlsson, R., editor, *Genomic Imprinting, Results and Problems in Cell Differentiation*, pages 41–71. Springer, Berlin, Heidelberg. ISBN 978-3-662-21956-0 978-3-540-69111-2.
- Hershkovitz, M. A. and Zimmer, E. A., 1997. On the evolutionary origins of the cacti. *Taxon*, 46(2):217–232.
- Hervé, M., 2015. Rvaidememoire: diverse basic statistical and graphical functions. *R package version 0.9-50*.

- Hillis, D. M. and Dixon, M. T., 1991. Ribosomal DNA: Molecular Evolution and Phylogenetic Inference. *The Quarterly Review of Biology*, 66(4):411–453.
- Hinz, H. L., Winston, R. L., and Schwarzländer, M., 2019. How safe is weed biological control? A global review of direct nontarget attack. *The Quarterly Review of Biology*, 94(1):1–27.
- Hoarau, G., Holla, S., Lescasse, R., Stam, W. T., and Olsen, J. L., 2002. Heteroplasmy and evidence for recombination in the mitochondrial control region of the flatfish *Platichthys flesus*. *Molecular Biology and Evolution*, 19(12):2261–2264. doi: 10.1093/oxfordjournals.molbev.a004049.
- Hodkinson, T. R., Chase, M. W., and Renvoize, S. A., 2002. Characterization of a genetic resource collection for *Miscanthus* (Saccharinae, Andropogoneae, Poaceae) using AFLP and ISSR PCR. *Annals of Botany*, 89(5):627–636.
- Hoffmann, J. H., 2003. Proceedings of the XI International Symposium on Biological Control of Weeds. In Cullen, J. M., Briese, D. T., Kriticos, D. J., Lonsdale, W. M., Morin, L., and Scott, J. K., editors, *Biotypes, hybrids and biological control: lessons from cochineal insects on Opuntia weeds*, pages 283–286. CSIRO Entomology, 2003.
- Hoffmann, J. H., 2004. Biotypes, hybrids and biological control: lessons from cochineal insects on *Opuntia* weeds. In *XI International Symposium on Biological Control of Weeds*, page 283, 2004.
- Hoffmann, J. H., Moran, V. C., and Zeller, D. A., 1998a. Evaluation of *Cactoblastis cactorum* (Lepidoptera: Phycitidae) as a Biological Control Agent of *Opuntia stricta* (Cactaceae) in the Kruger National Park, South Africa. *Biological Control*, 12(1):20–24. doi: 10.1006/bcon.1998.0608.
- Hoffmann, J. H., Impson, F. A. C., and Volchansky, C. R., 2002. Biological Control of Cactus Weeds: Implications of Hybridization between Control Agent Biotypes: Performance of Hybrid Cochineal Biotypes. *Journal of Applied Ecology*, 39(6):900–908. doi: 10.1046/j.1365-2664.2002.00766.x.
- Hoffmann, J., Moran, V., and Zeller, D., 1998b. Long-term population studies and the development of an integrated management programme for control of *Opuntia stricta* in Kruger National Park, South Africa. *Journal of Applied Ecology*, 35(1):156–160.
- Holland, B. R., Clarke, A. C., and Meudt, H. M., 2008. Optimizing automated AFLP scoring parameters to improve phylogenetic resolution. *Systematic Biology*, 57(3):347–366.
- Hoy, M. A., 2013. DNA, Gene Structure, and DNA Replication. In *Insect Molecular Genetics*, pages 3–36. Elsevier. ISBN 978-0-12-415874-0.
- Huelsenbeck, J. P. and Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17(8):754–5.
- Huelsenbeck, J. P. and Rannala, B., 2004. Frequentist properties of bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Systematic biology*, 53(6):904–913.
- Huelsenbeck, J. P., Ronquist, F., Nielsen, R., and Bollback, J. P., 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science*, 294(5550):2310–2314.
- Huelsenbeck, J. P., Larget, B., Miller, R. E., and Ronquist, F., 2002. Potential applications and pitfalls of Bayesian inference of phylogeny. *Systematic Biology*, 51(5):673–688.
- Hughes-Schrader, S., 1948. Cytology of Coccids (Coccodea-Homoptera). *Advances in Genetics*, 35(2): 127–203.

- Hulcr, J., Miller, S. E., Setliff, G. P., Darrow, K., Mueller, N. D., Hebert, P. D. N., and Weiblen, G. D., 2007. DNA Barcoding Confirms Polyphagy in a Generalist Moth, *Homona mermerodes* (Lepidoptera: Tortricidae). *Molecular Ecology Notes*, 7(4):549–557. doi: 10.1111/j.1471-8286.2007.01786.x.
- Humboldt, A., 1966. *Political essay on the kingdom of New Spain*. AMS Press, New York.
- Hundsdoerfer, A. and Wink, M., 2006. Incongruence of morphology and genetic markers in *Hyles tithymali* (Lepidoptera: Sphingidae) from the Canary Islands. *Journal of Zoological Systematics and Evolutionary Research*, 44(4):316–322.
- Hurst, L. D., 1991. The incidences and evolution of cytoplasmic male killers. *Proceedings of the Royal Society of London B*, 244(1310):91–99. doi: 10.1098/rspb.1991.0056.
- Huson, D. H. and Bryant, D., 2006. Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution*, 23(2):254–267.
- Iline, I. and Phillips, C., 2004. Allozyme markers to help define the South American origins of *Microctonus hyperodae* (Hymenoptera: Braconidae) established in New Zealand for biological control of Argentine stem weevil. *Bulletin of Entomological Research*, 94(3):229–234.
- Innocenti, P., Morrow, E. H., and Dowling, D. K., 2011. Experimental Evidence Supports a Sex-Specific Selective Sieve in Mitochondrial Genome Evolution. *Science*, 332(6031):845–848. doi: 10.1126/science.1201157.
- Invasive Plants and Animals Committee, 2017. Australian weeds strategy: 2017-2027. *Natural Resource Management Ministerial Council, Canberra*.
- Jaccard, P., 1908. Nouvelles recherches sur la distribution florale. *Bulletin de la Societe Vaudoise des Sciences Naturelles*, 44:223–70. doi: 10.5169/seals-268384.
- Jenuth, J. P., Peterson, A. C., Fu, K., and Shoubridge, E. A., 1996. Random Genetic Drift in the Female Germline Explains the Rapid Segregation of Mammalian Mitochondrial DNA. *Nature Genetics*, 14(2):146–151. doi: 10.1038/ng1096-146.
- Joger, U., Fritz, U., Guicking, D., Kalyabina-Hauf, S., Nagy, Z. T., and Wink, M., 2007. Phylogeography of western palaeartic reptiles-spatial and temporal speciation patterns. *Zoologischer Anzeiger-A Journal of Comparative Zoology*, 246(4):293–313.
- Jones, C., Edwards, K., Castaglione, S., Winfield, M., Sala, F., Van de Wiel, C., Bredemeijer, G., Vosman, B., Matthes, M., Daly, A., Brettschneider, R., Bettini, P., Buiatti, M., Maestri, E., Malcevski, A., Marmioli, N., Aert, R., Volckaert, G., Rueda, J., Linacero, R., Vazquez, A., and Karp, A., 1997. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Molecular Breeding*, 3(5):381–390.
- Jones, P., McConnachie, A., Holtkamp, R. H., Moore, K., Williams, A., and Day, M., 2018. Integrating results from host range, efficacy and fitness trials to prioritise biotypes of *Dactylopius tomentosus*. In *XV International Symposium on Biological Control of Weeds*, Hotel Bellevue, Engelberg, Switzerland. ISBCW.
- Jones, P. K., Holtkamp, R. H., Palmer, W. A., and Day, M. D., 2015. The Host Range of Three Biotypes of *Dactylopius tomentosus* (Lamarck) (Homoptera: Dactylopiidae) and Their Potential as Biological Control Agents of *Cylindropuntia* Spp. (Cactaceae) in Australia. *Biocontrol Science and Technology*, 25(6):613–628. doi: 10.1080/09583157.2014.999747.

- Jones, P., Holtkamp, R., and 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 Science and Technology*, 26(8):1033–1047.
- Jones, P., Oakey, J., Day, M., and 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., editors, *20th Australasian Weeds Conference 2016*. Weed Society of Western Australia, 2016b. ISBN 9781510834088.
- Jourdan, M., Thomann, T., Kriticos, D. J., Bon, M.-C., Sheppard, A., and Baker, G. H., 2019. Sourcing effective biological control agents of conical snails, *Cochlicella acuta*, in Europe and north Africa for release in southern Australia. *Biological Control*, 134:1–14.
- Kaplan, H., Wilson, J. R., Klein, H., Henderson, L., Zimmermann, H. G., Manyama, P., Ivey, P., Richardson, D. M., and Novoa, A., 2017. A proposed national strategic framework for the management of Cactaceae in South Africa. *Bothalia - African Biodiversity & Conservation*, 47(2):1–12.
- Katoh, K., Rozewicki, J., and Yamada, K. D., 2017. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*. doi: 10.1093/bib/bbx108.
- Keane, R. M. and Crawley, M. J., 2002. Exotic plant invasions and the enemy release hypothesis. *Trends in Ecology & Evolution*, 17(4):164–170.
- Kerrigan, J., Smith, M. T., Rogers, J. D., Douhan, G. W., et al., 2003. *Ascobotryozyma cognata* sp. nov., a new ascomycetous yeast associated with nematodes from wood-boring beetle galleries. *Mycological Research*, 107(9):1110–1120.
- Klarica, J., Kloss-Brandstätter, A., Traugott, M., and Juen, A., 2012. Comparing four mitochondrial genes in earthworms - implications for identification, phylogenetics, and discovery of cryptic species. *Soil Biology and Biochemistry*, 45:23–30.
- 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. *African Entomology*, 19(2):515–549. doi: 10.4001/003.019.0214.
- Klein, H., 2002. *Weeds Biocontrol*. PPRI, Pretoria. ISBN 1-86849-111-8.
- Kmiec, B., Woloszynska, M., and Janska, H., 2006. Heteroplasmy as a common state of mitochondrial genetic information in plants and animals. *Current Genetics*, 50(3):149–159. doi: 10.1007/s00294-006-0082-1.
- Knowlton, N. and Weigt, L. A., 1998. New dates and new rates for divergence across the Isthmus of Panama. *Proceedings of the Royal Society B: Biological Sciences*, 265(1412):2257. doi: 10.1098/rspb.1998.0568.
- Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., and Mayrose, I., 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, 15(5):1179–1191. doi: 10.1111/1755-0998.12387.
- Kotzé, J. D. F., Beukes, B. H., van den Berg, E. C., and Newby, T. S., 2010. National Invasive Alien Plant Survey. Technical report, Agricultural Research Council: Institute for Soil, Climate and Water, Pretoria.
- Kriticos, D., 2003. The roles of ecological models in evaluating weed biological control agents and projects. In Spafford-Jacob, H. and Briese, D., editors, *Improving the Selection, Testing and Evaluation of Weed Biological Control Agents*, pages 69–74. Cooperative Research Centre for Australian Weed Management, Adelaide, Australia.

- Kueffer, C. and Kull, C. A., 2017. Non-native Species and the Aesthetics of Nature. In Vilà, M. and Hulme, P. E., editors, *Impact of Biological Invasions on Ecosystem Services*, pages 311–324. Springer International Publishing, Cham. doi: 10.1007/978/3/319/45121/3/20.
- Kumar, S., Stecher, G., and Tamura, K., 2016. Mega7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular biology and evolution*, 33(7):1870–1874.
- Kvist, L., Martens, J., Nazarenko, A. A., and Orell, M., 2003. Paternal leakage of mitochondrial DNA in the great tit (*Parus major*). *Molecular Biology and Evolution*, 20(2):243–247. doi: 10.1093/molbev/msg025.
- Labra, M., Grassi, F., Bardini, M., Imazio, S., Guiggi, A., Citterio, S., Banfi, E., and Sgorbati, S., 2003. Genetic Relationships in *Opuntia* Mill. Genus (Cactaceae) Detected by Molecular Marker. *Plant Science*, 165(5):1129–1136. doi: 10.1016/S0168-9452(03)00321-2.
- Ladoukakis, E. D. and Zouros, E., 2017. Evolution and Inheritance of Animal Mitochondrial DNA: Rules and Exceptions. *Journal of Biological Research-Thessaloniki*, 24(1). doi: 10.1186/s40709-017-0060-4.
- Lawrence, E., 2011. *Henderson's Dictionary of Biology*. Pearson, 15th edition.
- Le Maitre, D. C., Versfeld, D. B., Chapman, R. A., Maitre, L., C, D., Versfeld, D. B., and Chapman, R. A., 2000. Impact of Invading Alien Plants on Surface Water Resources in South Africa: A Preliminary Assessment. *Water SA*, 26(3):397–408.
- Le Maitre, D. C., Kotzee, I. M., O'Farrell, P. J., O'Farrell, P. J., and O'Farrell, P. J., 2014. Impacts of Land-Cover Change on the Water Flow Regulation Ecosystem Service: Invasive Alien Plants, Fire and Their Policy Implications. *Land Use Policy*, 36:171–181. doi: 10.1016/j.landusepol.2013.07.007.
- Leigh, J. W. and Bryant, D., 2015. popart: full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, 6(9):1110–1116.
- Leite, L. A. R., 2012. Mitochondrial pseudogenes in insect DNA barcoding: differing points of view on the same issue. *Biota Neotropica*, 12(3):301–308. doi: 10.1590/S1676-06032012000300029.
- Lewis, W. M., 1985. Nutrient Scarcity as an Evolutionary Cause of Haploidy. *The American Naturalist*, 125(5):692–701. doi: 10.1086/284372.
- Li, L., Xu, X., Wu, P., Zhang, G., and Zhang, X., 2018. Establishment of Molecular Identity Cards for *Cucumis melo* Cultivars Using SSR Markers. *HortScience*, 53(2):138–143.
- Li, Y.-L. and Liu, J.-X., 2018. StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. *Molecular Ecology Resources*, 18(1):176–177.
- Lin, X., Stur, E., and Ekrem, T., 2015. Exploring Genetic Divergence in a Species-Rich Insect Genus Using 2790 DNA Barcodes. *Public Library of Science One*, 10(9):e0138993. doi: 10.1371/journal.pone.0138993.
- Liu, H. and Stiling, P., 2006. Testing the enemy release hypothesis: a review and meta-analysis. *Biological Invasions*, 8(7):1535–1545.
- Lloyd, S. and Reeves, A., 2014. *Situation statement on opuntiod cacti (Austrocylindropuntia spp., Cylindropuntia spp., and Opuntia spp.) in Western Australia*. Invasive Species Program, Department of Agriculture and Food, Western Australia.
- Lockwood, J. L., Hoopes, M. F., and Marchetti, M. P., 2013. *Invasion Ecology*. John Wiley & Sons, West Sussex, 2nd edition. ISBN 1118570820.

- Lopez, J. V., Yuhki, N., Masuda, R., Modi, W., and O'Brien, S. J., 1994. Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *Journal of Molecular Evolution*, 39(2):174–190. doi: 10.1007/bf00163806.
- Lotter, W. and Hoffmann, J., 1998. An integrated management plan for the control of *Opuntia stricta* (Cactaceae) in the Kruger National Park, South Africa. *Koedoe*, 41(1):63–68.
- Lounsbury, C., 1915. Plant killing insects: the Indian cochineal. *Agricultural Journal of South Africa*, 1:537–543.
- Lunt, D. H., Zhang, D.-X., Szymura, J. M., and Hewlitt, O. M., 1996. The insect cytochrome oxidase I gene: evolutionary patterns and conserved primers for phylogenetic studies. *Insect Molecular Biology*, 5(3):153–165. doi: 10.1111/j.1365-2583.1996.tb00049.x.
- Luo, J., Yan, D., Zhang, D., Han, Y., Dong, X., Yang, Y., Deng, K., and Xiao, X., 2011. Application of 12S rRNA Barcodes for the Identification of Animal-Derived Drugs. *Journal of Pharmacy & Pharmaceutical Sciences*, 14(3):358–367.
- Lv, J., Wu, S., Zhang, Y., Chen, Y., Feng, C., Yuan, X., Jia, G., Deng, J., Wang, C., Wang, Q., Mei, L., and Lin, X., 2014. Assessment of four DNA fragments (COI, 16S rDNA, ITS2, 12S rDNA) for species identification of the Ixodida (Acari: Ixodida). *Parasites & Vectors*, 7(1):93.
- Mable, B. K. and Otto, S. P., 1998. The Evolution of Life Cycles with Haploid and Diploid Phases. *BioEssays*, 20(6):453–462. doi: 10.1002/(SICI)1521-1878(199806)20:6<453::AID-BIES3>3.0.CO;2-N.
- Machkour-M'Rabet, S., Hénaut, Y., Charreau, P., Gevrey, M., Winterton, P., and Legal, L., 2009a. Between introgression events and fragmentation, islands are the last refuge for the American crocodile in Caribbean Mexico. *Marine Biology*, 156(6):1321–1333.
- Machkour-M'Rabet, S., Hénaut, Y., Dor, A., Pérez-Lachaud, G., Péliissier, C., Gers, C., and Legal, L., 2009b. ISSR (inter simple sequence repeats) as molecular markers to study genetic diversity in tarantulas (Araneae, Mygalomorphae). *Journal of Arachnology*, 37(1):10–14.
- Maddison, W. P., 1997. Gene Trees in Species Trees. *Systematic Biology*, 46(3):523–536. doi: 10.1093/sysbio/46.3.523.
- Majure, L. C., Puente, R., Griffith, M. P., Judd, W. S., Soltis, P. S., and Soltis, D. E., 2012. Phylogeny of *Opuntia* s.s. (Cactaceae): Clade Delineation, Geographic Origins, and Reticulate Evolution. *American Journal of Botany*, 99(5):847–864. doi: 10.3732/ajb.1100375.
- Mann, J., 1969. Cactus-feeding Insects and Mites. *Bulletin of the United States National Museum*, 256 (256):1–158. doi: 10.5479/si.03629236.256.1.
- Mann, J., 1970. *Cacti naturalised in Australia and their control*. Brisbane (Australia) Department of Lands, Queensland.
- Mathenge, C. W., Holford, P., Hoffmann, J. H., Spooner-Hart, R., Beattie, G. a. C., and Zimmermann, H. G., 2009a. The Biology of *Dactylopius tomentosus* (Hemiptera: Dactylopiidae). *Bulletin of Entomological Research*, 99(6):551–559. doi: 10.1017/S0007485308006597.
- Mathenge, C. W., Holford, P., Hoffmann, J. H., Zimmermann, H. G., Spooner-Hart, R., and Beattie, G. A. C., 2009b. Distinguishing Suitable Biotypes of *Dactylopius tomentosus* (Hemiptera: Dactylopiidae) for Biological Control of *Cylindropuntia fulgida* var. *fulgida* (Caryophyllales: Cactaceae) in South Africa. *Bulletin of Entomological Research*, 99(06):619. doi: 10.1017/S0007485309006671.

- Mathenge, C. W., Holford, P., Hoffmann, J. H., Zimmermann, H. G., Spooner-Hart, R., and Beattie, G. A. C., 2010a. Determination of Biotypes of *Dactylopius tomentosus* (Hemiptera: Dactylopiidae) and Insights into the Taxonomic Relationships of Their Hosts, *Cylindropuntia* Spp. *Bulletin of Entomological Research*, 100(03):347–358. doi: 10.1017/S0007485309990496.
- Mathenge, C. W., Holford, P., Hoffmann, J. H., Zimmermann, H. G., Spooner-Hart, R. N., and Beattie, G. A. C., 2010b. Hybridization between *Dactylopius tomentosus* (Hemiptera: Dactylopiidae) Biotypes and Its Effects on Host Specificity. *Bulletin of Entomological Research*, 100(03):331–338. doi: 10.1017/S0007485309990344.
- Mathenge, C. W., Riegler, M., Beattie, G. A. C., Spooner-Hart, R. N., and Holford, P., 2015. Genetic variation amongst biotypes of *Dactylopius tomentosus*. *Insect Science*, 22(3):360–374. doi: 10.1111/1744-7917.12120.
- Mauseth, J. D., 1990. Continental drift, climate and the evolution of cacti. *Cactus and Succulent Journal*, 62(6):302–308.
- Maxwell, F. G. and Jennings, P. R., 1980. *Breeding Plants Resistant to Insects*. Number 631.53 B832. John Wiley.
- Mayr, E., 1963. *Animal species and evolution*. Harvard University Press.
- Mayr, E., 1970. *Populations, species, and evolution: an abridgment of animal species and evolution*, volume 19. Harvard University Press.
- Mazzeo, G., Nucifora, S., Russo, A., and Suma, P., 2019. *Dactylopius opuntiae*, a new prickly pear cactus pest in the Mediterranean: an overview. *Entomologia Experimentalis et Applicata*, 167:59–72.
- McAndrew, B. and Majumdar, K., 1983. Tilapia stock identification using electrophoretic markers. *Aquaculture*, 30(1-4):249–261. doi: 10.1016/0044-8486(83)90167-9.
- McConnachie, A. J., Strathie, L. W., Mersie, W., Gebrehiwot, L., Zewdie, K., Abdurehim, A., Abrha, B., Araya, T., Asaregew, F., Assefa, F., Gebre-Tsadik, R., Nigatu, L., Tadesse, B., and Tana, T., 2011. Current and Potential Geographical Distribution of the Invasive Plant *Parthenium hysterophorus* (Asteraceae) in Eastern and Southern Africa: Distribution of *Parthenium hysterophorus* in Eastern and Southern Africa. *Weed Research*, 51(1):71–84. doi: 10.1111/j.1365-3180.2010.00820.x.
- McEvoy, P. B., 1996. Host specificity and biological pest control. *BioScience*, 46(6):401–405.
- McEvoy, P. B., 2002. Insect-plant interactions on a planet of weeds. *Entomologia Experimentalis et Applicata*, 104(1):165–179.
- McEvoy, P. B. and Coombs, E. M., 1999. Biological control of plant invaders: regional patterns, field experiments, and structured population models. *Ecological Applications*, 9(2):387–401.
- McFadyen, R. E. C., 1998. Biological control of weeds. *Annual Review of Entomology*, 43(1):369–393.
- McFadyen, R. and Marohasy, J., 1990. A leaf feeding moth, *Euclasta whalleyi* [Lep.: Pyralidae] for the biological control of *Cryptostegia grandiflora* [Asclepiadaceae] in Queensland, Australia. *Entomophaga*, 35(3):431–435.
- McGregor, C., Lambert, C., Greyling, M., Louw, J., and Warnich, L., 2000. A comparative assessment of DNA fingerprinting techniques (RAPD, ISSR, AFLP and SSR) in tetraploid potato (*Solanum tuberosum* L.) germplasm. *Euphytica*, 113(2):135–144.

- Medrano, M., López-Perea, E., and Herrera, C. M., 2014. Population genetics methods applied to a species delimitation problem: Endemic trumpet daffodils (*Narcissus* section *Pseudonarcissi*) from the Southern Iberian Peninsula. *International Journal of Plant Sciences*, 175(5):501–517.
- Meier, R., Shiyang, K., Vaidya, G., and Ng, P. K. L., 2006. DNA Barcoding and Taxonomy in Diptera: A Tale of High Intraspecific Variability and Low Identification Success. *Systematic Biology*, 55(5): 715–728. doi: 10.1080/10635150600969864.
- Meiklejohn, K. A., Wallman, J. F., and Dowton, M., 2013. DNA barcoding identifies all immature life stages of a forensically important flesh fly (Diptera: Sarcophagidae). *Journal of Forensic Sciences*, 58(1):184–187. doi: 10.1111/j.1556-4029.2012.02220.x.
- Melters, D. P., Paliulis, L. V., Korf, I. F., and Chan, S. W. L., 2012. Holocentric chromosomes: convergent evolution, meiotic adaptations, and genomic analysis. *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology*, 20(5):579–593. doi: 10.1007/s10577-012-9292-1.
- Meudt, H. M. and Clarke, A. C., 2007. Almost forgotten or latest practice? AFLP applications, analyses and advances. *Trends in Plant Science*, 12(3):106–117.
- Meusel, M. S. and Moritz, R. F., 1993. Transfer of Paternal Mitochondrial DNA during Fertilization of Honeybee (*Apis mellifera* L.) Eggs. *Current Genetics*, 24(6):539–543.
- Meyer, M. C. and Beezley, W. H., 2000. *The Oxford history of Mexico*. Oxford University Press, USA.
- Miller, D. R. and Kosztarab, M., 1979. Recent advances in the study of scale insects. *Annual Review of Entomology*, 24(1):1–27.
- Miller, M., Pfeiffer, W., and Schwartz, T., 2010. Creating the Cipres Science gateway for inference of large phylogenetic trees. In *Proceedings of the Gateway Computing Environments Workshop (GCE)*, pages 1–8, New Orleans, LA.
- Milton, S. J., 2004. Grasses as Invasive Alien Plants in South Africa: Working for Water. *South African Journal of Science*, 100(1-2):69–75.
- Montes, C., Cardona, A., Jaramillo, C., Pardo, A., Silva, J., Valencia, V., Ayala, C., Pérez-Angel, L., Rodríguez-Parra, L., Ramirez, V., *et al.*, 2015. Middle Miocene closure of the Central American seaway. *Science*, 348(6231):226–229.
- Moran, V. C. and Cabby, B. S., 1979. On the Life-History and Fecundity of the Cochineal Insect, *Dactylopius austrinus* De Lotto (Homoptera: Dactylopiidae), a Biological Control Agent for the Cactus *Opuntia awantiaca*. *Bulletin of Entomological Research*, 69(04):629. doi: 10.1017/S0007485300020174.
- Moran, V. C. and Zimmermann, H. G., 1991a. Biological Control of Jointed Cactus, *Opuntia aurantiaca* (Cactaceae), in South Africa. *Agriculture, Ecosystems & Environment*, 37(1-3):5–27. doi: 10.1016/0167-8809(91)90136-L.
- Moran, V. C. and Zimmermann, H. G., 1991b. Biological Control of Cactus Weeds of Minor Importance in South Africa. *Agriculture, Ecosystems & Environment*, 37(1):37–55. doi: 10.1016/0167-8809(91)90138-N.
- Moran, V. C., Hoffmann, J. H., and Zimmermann, H. G., 2005. Biological Control of Invasive Alien Plants in South Africa: Necessity, Circumspection, and Success. *Frontiers in Ecology and the Environment*, 3(2):77–83. doi: 10.2307/3868513.

- Moretti, S., Wilm, A., Higgins, D. G., Xenarios, I., and Notredame, C., 2008. R-Coffee: a web server for accurately aligning noncoding RNA sequences. *Nucleic Acids Research*, 36(suppl\_2):W10–W13.
- Morrison, D. A., 2006. L. A. S. Johnson Review No. 8. Multiple sequence alignment for phylogenetic purposes. *Australian Systematic Botany*, 19(6):479. doi: 10.1071/SB06020.
- Moulton, M. J., Song, H., and Whiting, M. F., 2010. Assessing the effects of primer specificity on eliminating numt coamplification in DNA barcoding: a case study from Orthoptera (Arthropoda: Insecta). *Molecular Ecology Resources*, 10(4):615–627. doi: 10.1111/j.1755-0998.2009.02823.x.
- Naeemullah, M., Sharma, P. N., Tufail, M., Mori, N., Matsumura, M., Takeda, M., and Nakamura, C., 2009. Characterization of brown planthopper strains based on their differential responses to introgressed resistance genes and on mitochondrial DNA polymorphism. *Applied Entomology and Zoology*, 44(3):475–483.
- Nagoshi, R. N., Brambila, J., and Meagher, R. L., 2011. Use of DNA Barcodes to Identify Invasive Armyworm *Spodoptera* Species in Florida. *Journal of Insect Science*, 11(154):1–11. doi: 10.1673/031.011.15401.
- Nagy, Z. T., Glaw, F., Andreone, F., Wink, M., and Vences, M., 2007. Species boundaries in Malagasy snakes of the genus *Madagascarophis* (Serpentes: Colubridae sensu lato) assessed by nuclear and mitochondrial markers. *Organisms Diversity & Evolution*, 7(3):241–251.
- Nanney, D. L., 1982. Genes and Phenotypes in *Tetrahymena*. *BioScience*, 32(10):783–788. doi: 10.2307/1308971.
- Nelson-Rees, W. A., Hoy, M. A., and Roush, R. T., 1980. Heterochromatinization, Chromatin Elimination and Haploidization in the Parahaploid Mite *Metaseiulus occidentalis* (Nesbitt) (Acarina: Phytoseiidae). *Chromosoma*, 77(3):263–276.
- Nelson-Rees, W. A., 1962. The Effects of Radiation Damaged Heterochromatic Chromosomes on Male Fertility in the Mealy Bug, *Planococcus citri* (Risso). *Genetics*, 47(6):661–683.
- Nesi, N., Nakouné, E., Cruaud, C., and Hassanin, A., 2011. DNA barcoding of African fruit bats (Mammalia, Pteropodidae). The mitochondrial genome does not provide a reliable discrimination between *Epomophorus gambianus* and *Micropteropus pusillus*. *Comptes Rendus Biologies*, 334(7): 544–554. doi: 10.1016/J.CRVI.2011.05.003.
- Ng, W. L. and Tan, S. G., 2015. Inter-Simple Sequence Repeat (ISSR) Markers: Are We Doing It Right? *The Academy of Sciences Malaysia Science Journal*, 9(1):11.
- Nichols, R., 2001. Gene trees and species trees are not the same. *Trends in Ecology & Evolution*, 16(7):358–364. doi: 10.1016/S0169-5347(01)02203-0.
- Nobel, P. S., 2002. *Cacti : Biology and Uses*. University of California Press. ISBN 0520231570.
- Normark, B. B., 2004a. Haplodiploidy as an outcome of coevolution between male-killing cytoplasmic elements and their hosts. *Evolution*, 58(4):790–798. doi: 10.1111/j.0014-3820.2004.tb00412.x.
- Normark, B. B., 2004b. The Strange Case of the Armored Scale Insect and Its Bacteriome. *Public Library of Science Biology*, 2(3):e43. doi: 10.1371/journal.pbio.0020043.
- Normark, B. B., 2006. Perspective: maternal kin groups and the origins of asymmetric genetic systems: genomic imprinting, haplodiploidy, and parthenogenesis. *Evolution*, 60(4):631–642. doi: 10.1111/j.0014-3820.2006.tb01145.x.

- Notredame, C., Higgins, D. G., and Heringa, J., 2000. T-coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology*, 302(1):205–217.
- Novoa, A., Le Roux, J. J., Robertson, M. P., Wilson, J. R. U., and Richardson, D. M., 2015. Introduced and Invasive Cactus Species: A Global Review. *Annals of Botany Plants*, 7(0):plu078–plu078. doi: 10.1093/aobpla/plu078.
- Novoa, A., Kaplan, H., Wilson, J. R., and Richardson, D. M., 2016. Resolving a prickly situation: Involving stakeholders in invasive cactus management in South Africa. *Environmental Management*, 57(5):998–1008.
- Novoa, A., Flepu, V., and Boatwright, J. S., 2019. Is spinelessness a stable character in cactus pear cultivars? Implications for invasiveness. *Journal of Arid Environments*, 160:11–16.
- Nur, U., 1967. Reversal of Heterochromatization and the Activity of the Paternal Chromosome Set in the Male Mealy Bug. *Genetics*, 56(3):375–389.
- Nur, U., 1980. Evolution of unusual chromosome systems in scale insects (Coccoidea: Homoptera). In Hewitt, G. M. and Ashburner, M., editors, *Insect Cytogenetics*, pages 97–118. Blackwell, Oxford.
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., and Wagner, H. *Vegan: Community Ecology Package*, 2018. URL <https://CRAN.R-project.org/package=vegan>. R package version 2.5-3.
- Olckers, T., Zimmermann, H., and Hoffmann, J., 1995. Interpreting ambiguous results of host-specificity tests in biological control of weeds: assessment of two *Leptinotarsa* species (Chrysomelidae) for the control of *Solanum elaeagnifolium* (Solanaceae) in South Africa. *Biological Control*, 5(3):336–344.
- Otto, S. P. and Gerstein, A. C., 2008. The Evolution of Haploidy and Diploidy. *Current Biology*, 18(24):R1121–R1124. doi: 10.1016/j.cub.2008.09.039.
- Painter, R. H., 1941. The economic value and biologic significance of insect resistance in plants. *Journal of Economic Entomology*, 34(3):358–367.
- Pamilo, P., Viljakainen, L., and Vihavainen, A., 2007. Exceptionally High Density of NUMTs in the Honeybee Genome. *Molecular Biology and Evolution*, 24(6):1340–1346. doi: 10.1093/molbev/msm055.
- Paquin, P. and Hedin, M., 2011. The power and perils of molecular taxonomy: a case study of eyeless and endangered *Cicurina* (Araneae: Dictynidae) from Texas caves. *Molecular Ecology*, 13(10):3239–3255. doi: 10.1111/J.1365-294X.2004.02296.X.
- Paradis, E., Claude, J., and Strimmer, K., 2004. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, 20(2):289–290. doi: 10.1093/bioinformatics/btg412.
- Park, D.-S., Suh, S.-J., Oh, H.-W., and Hebert, P. D. N., 2010. Recovery of the Mitochondrial COI Barcode Region in Diverse Hexapoda through tRNA-Based Primers. *BMC Genomics*, 11:7. doi: 10.1186/1471-2164-11-423.
- Paterson, I. D. and Zachariades, C., 2013. ISSRs indicate that *Chromolaena odorata* invading southern Africa originates in Jamaica or Cuba. *Biological Control*, 66(2):132–139.
- Paterson, I. D., Downie, D. A., and Hill, M. P., 2009. Using molecular methods to determine the origin of weed populations of *Pereskia aculeata* in South Africa and its relevance to biological control. *Biological Control*, 48(1):84–91. doi: 10.1016/j.biocontrol.2008.09.012.

- Paterson, I. D., Mangan, R., Downie, D. A., Coetzee, J. A., Hill, M. P., Burke, A. M., Downey, P. O., Henry, T. J., and Compton, S. G., 2016. Two in One: Cryptic Species Discovered in Biological Control Agent Populations Using Molecular Data and Crossbreeding Experiments. *Ecology and Evolution*, 6(17):6139–6150. doi: 10.1002/ece3.2297.
- Paterson, I., Vitorino, M., de Cristo, S., Martin, G., and Hill, M., 2014. Prioritisation of potential agents for the biological control of the invasive alien weed, *Pereskia aculeata* (Cactaceae), in South Africa. *Biocontrol Science and Technology*, 24(4):407–425.
- Paterson, I., Manheimmer, C., and Zimmermann, H., 2019. Prospects for biological control of cactus weeds in Namibia. *Biocontrol Science and Technology*, 29(4):393–399.
- Paterson, I. D., Hoffmann, J. H., Klein, H., Mathenge, C. W., Naser, S., and Zimmermann, H. G., 2011. Biological Control of Cactaceae in South Africa. *African Entomology*, 19(SP):230–246. doi: 10.4001/003.019.0221.
- Paynter, Q., Fowler, S. V., Gourlay, A. H., Peterson, P. G., Smith, L. A., and Winks, C. J., 2015. Relative performance on test and target plants in laboratory tests predicts the risk of non-target attack in the field for arthropod weed biocontrol agents. *Biological Control*, 80:133–142.
- Paynter, Q., Fowler, S. V., and Groenteman, R., 2018. Making weed biological control predictable, safer and more effective: perspectives from New Zealand. *BioControl*, 63(3):427–436.
- Pejchar, L. and Mooney, H. A., 2009. Invasive species, ecosystem services and human well-being. *Trends in Ecology & Evolution*, 24(9):497–504. doi: 10.1016/j.tree.2009.03.016.
- Penny, D., 2005. Evolutionary biology: Relativity for molecular clocks. *Nature*, 436(7048):183–184. doi: 10.1038/436183a.
- Pérez-Guerra, G. and 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.
- Perring, T. M., 2001. The *Bemisia tabaci* species complex. *Crop Protection*, 20(9):725–737.
- Petit, R. J. and Excoffier, L., 2009. Gene flow and species delimitation. *Trends in Ecology & Evolution*, 24(7):386–393. doi: 10.1016/J.TREE.2009.02.011.
- Philippe, H., Brinkmann, H., Lavrov, D. V., Littlewood, D. T. J., Manuel, M., Wörheide, G., and Baurain, D., 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. *Public Library of Science Biology*, 9(3):e1000602.
- Piganeau, G., Gardner, M., and Eyre-Walker, A., 2004. A Broad Survey of Recombination in Animal Mitochondria. *Molecular Biology and Evolution*, 21(12):2319–2325. doi: 10.1093/molbev/msh244.
- Pompanon, F., Bonin, A., Bellemain, E., and Taberlet, P., 2005. Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics*, 6(11):847.
- Pook, C. and McEwing, R., 2005. Mitochondrial DNA sequences from dried snake venom: a DNA barcoding approach to the identification of venom samples. *Toxicon*, 46(7):711–715.
- Porter, D. R., Burd, J. D., Shufran, K. A., Webster, J. A., and Teetes, G. L., 1997. Greenbug (Homoptera: Aphididae) biotypes: selected by resistant cultivars or preadapted opportunists? *Journal of Economic Entomology*, 90(5):1055–1065.

- Portillo, L., 2007. Biogeography of Dactylopiidae and human factor. In *VI International Congress on Cactus Pear and Cochineal 811*, pages 235–240, 2007.
- Portillo, M. and Viguera, A., 2006. A Review on the cochineal species in Mexico, hosts and natural enemies. *Acta Horticulturae*, (728):249–256. doi: 10.17660/ActaHortic.2006.728.35.
- Prince, L., 2009. The relationship of *Monardella viminea* to closely related taxa based on analyses of ISSRs. *USFWS Report*, page P0750003.
- Printz, Y., 1937. Contribution to the question of the changes in the virulence of *Phylloxera* of different biotypes. *Plant Protection Leningrad*, 12:137–142.
- Pritchard, J. K., Stephens, M., and Donnelly, P., 2000. Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, 155(2):945–959.
- Puechmaile, S. J., 2016. The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Molecular Ecology Resources*, 16(3):608–627.
- Pulquério, M. J. F. and Nichols, R. A., 2007. Dates from the Molecular Clock: How Wrong Can We Be? *Trends in Ecology & Evolution*, 22(4):180–184. doi: 10.1016/j.tree.2006.11.013.
- R Core Team. R: A language and environment for statistical computing, 2013. URL <http://www.R-project.org/>.
- Rabitsch, W., Essl, F., and Schindler, S., 2017. The Rise of Non-native Vectors and Reservoirs of Human Diseases. In Vilà, M. and Hulme, P. E., editors, *Impact of Biological Invasions on Ecosystem Services*, pages 263–275. Springer International Publishing, Cham. doi: 10.1007/978/3/319/45121/3/17.
- Raghu, S. and Walton, C., 2007. Understanding the ghost of *Cactoblastis* past: historical clarifications on a poster child of classical biological control. *BioScience*, 57(8):699–705.
- Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A., 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, 67(5):901–904.
- Ramírez-Puebla, S. T., Rosenblueth, M., Chávez-Moreno, C. K., Catanho Pereira de Lyra, M. C., Tecante, A., and Martínez-Romero, E., 2010. Molecular Phylogeny of the Genus *Dactylopius* (Hemiptera: Dactylopiidae) and Identification of the Symbiotic Bacteria. *Environmental Entomology*, 39(4): 1178–1183. doi: 10.1603/EN10037.
- Raupach, M. J., Astrin, J. J., Hannig, K., Peters, M. K., Stoeckle, M. Y., and Wägele, J.-W., 2010. Molecular species identification of Central European ground beetles (Coleoptera: Carabidae) using nuclear rDNA expansion segments and DNA barcodes. *Frontiers in Zoology*, 7(1):26. doi: 10.1186/1742-9994-7-26.
- Raupach, M. J., Hendrich, L., Kuchler, S. M., Deister, F., Morinière, J., and Gossner, M. M., 2014. Building-Up of a DNA Barcode Library for True Bugs (Insecta: Hemiptera: Heteroptera) of Germany Reveals Taxonomic Uncertainties and Surprises. *Public Library of Science One*, 9(9):e106940. doi: 10.1371/journal.pone.0106940.
- Razowski, J., Pelz, V., and Tarcz, S., 2017. Uncovering the hidden biodiversity of natural history collections: Insights from DNA barcoding and morphological characters of the Neotropical genus *Orthocomotis* Dognin (Lepidoptera: Tortricidae). *Zootaxa*, 4250(6):541. doi: 10.11646/zootaxa.4250.6.3.

- Reddy, M. P., Sarla, N., and Siddiq, E. A., 2002. Inter Simple Sequence Repeat (ISSR) Polymorphism and Its Application in Plant Breeding. *Euphytica*, 128(1):9–17.
- Richardson, D. M. and van Wilgen, B. W., 2004. Invasive Alien Plants in South Africa: How Well Do We Understand the Ecological Impacts? *South African Journal of Science*, 100:45–52.
- Richly, E. and Leister, D., 2004. NUMTs in Sequenced Eukaryotic Genomes. *Molecular Biology and Evolution*, 21(6):1081–1084. doi: 10.1093/molbev/msh110.
- Rigaud, T., 1997. Inherited microorganisms and sex determination of arthropod hosts. In O’Neill, S. L., Hoffmann, A. A., and Werren, J. H., editors, *Influential passengers: Inherited Microorganisms and Arthropod Reproduction*, pages 81–101. Oxford University Press.
- Robertson, M. P., Kriticos, D. J., and Zachariades, C., 2008. Climate matching techniques to narrow the search for biological control agents. *Biological Control*, 46(3):442–452.
- Rodríguez, L. C., Méndez, M. A., and Niemeyer, H. M., 2001. Direction of Dispersion of Cochineal (*Dactylopius coccus* Costa) within the Americas. *Antiquity*, 75(287):73–77. doi: 10.1017/S0003598X00052741.
- Rokas, A., Ladoukakis, E., and Zouros, E., 2003. Animal Mitochondrial DNA Recombination Revisited. *Trends in Ecology & Evolution*, 18(8):411–417. doi: 10.1016/S0169-5347(03)00125-3.
- Rollo, F., Ubaldi, M., Ermini, L., and Marota, I., 2002. Ötzi’s Last Meals: DNA Analysis of the Intestinal Content of the Neolithic Glacier Mummy from the Alps. *Proceedings of the National Academy of Sciences*, 99(20):12594–12599. doi: 10.1073/pnas.192184599.
- Rosenblueth, M., Martínez-Romero, J., and Ramírez-Puebla, S. T., 2018. Endosymbiotic microorganisms of scale insects. *Revista Especializada en Ciencias Químico-Biológicas*, 21(1):53–69. doi: 10.1016/j.recqb.2017.08.006.
- Ross, L., Langenhof, M. B. W., Pen, I., Beukeboom, L. W., West, S. A., and Shuker, D. M., 2010a. Sex Allocation in a Species with Paternal Genome Elimination: The Roles of Crowding and Female Age in the Mealybug *Planococcus citri*. *Evolutionary Ecology Research*, 12(1):89–104.
- Ross, L., Pen, I., and Shuker, D. M., 2010b. Genomic Conflict in Scale Insects: The Causes and Consequences of Bizarre Genetic Systems. *Biological Reviews of the Cambridge Philosophical Society*, 85(4):807–828. doi: 10.1111/j.1469-185X.2010.00127.x.
- Ross, L., Shuker, D. M., Normark, B. B., and Pen, I., 2012. The Role of Endosymbionts in the Evolution of Haploid-Male Genetic Systems in Scale Insects (Coccoidea). *Ecology and Evolution*, 2(5):1071–1081. doi: 10.1002/ece3.222.
- Roux, O., Gevrey, M., Arvanitakis, L., Gers, C., Bordat, D., and Legal, L., 2007. ISSR-PCR: Tool for discrimination and genetic structure analysis of *Plutella xylostella* populations native to different geographical areas. *Molecular Phylogenetics and Evolution*, 43(1):240–250.
- RStudio, Inc. *Easy web applications in R.*, 2013. URL: <http://www.rstudio.com/shiny/>.
- Rubinoff, D., 2006. Essays: Utility of Mitochondrial DNA Barcodes in Species Conservation: DNA Barcodes and Conservation. *Conservation Biology*, 20(4):1026–1033. doi: 10.1111/j.1523-1739.2006.00372.x.
- Rubinoff, D. and Sperling, F. A. H., 2004. Mitochondrial DNA Sequence, Morphology and Ecology Yield Contrasting Conservation Implications for Two Threatened Buckmoths (Hemileuca: Saturniidae). *Biological Conservation*, 118(3):341–351. doi: 10.1016/j.biocon.2003.09.013.

- Rubinoff, D., Cameron, S., and Will, K., 2006. A Genomic Perspective on the Shortcomings of Mitochondrial DNA for Barcoding Identification. *Journal of Heredity*, 97(6):581–594. doi: 10.1093/jhered/esl036.
- Rule, N. F. and Hoffmann, J., 2018. The performance of *Dactylopius opuntiae* as a biological control agent on two invasive *Opuntia* cactus species in South Africa. *Biological Control*, 119:7–11.
- Sabour, M., 1972. RNA synthesis and heterochromatization in early development of a mealybug. *Genetics*, 70(2):291–298.
- Saha, D., Ranjan, S. K., Mallick, C. B., Vidyarthi, A. S., and Ramani, R., 2011. Genetic diversity in lac resin-secreting insects belonging to *Kerria* spp., as revealed through ISSR markers. *Biochemical Systematics and Ecology*, 39(2):112–120.
- Savolainen, V., Cowan, R. S., Vogler, A. P., Roderick, G. K., and Lane, R., 2005. Towards Writing the Encyclopaedia of Life: An Introduction to DNA Barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1462):1805–1811. doi: 10.1098/rstb.2005.1730.
- Schaffner, U., 2001. Host range testing of insects for biological weed control: How can it be better interpreted? data on the host range of biocontrol candidates are particularly relevant in assessing potential detrimental effects to nontarget organisms. *BioScience*, 51(11):951–959.
- Schaffner, U., Smith, L., and Cristofaro, M., 2018. A review of open-field host range testing to evaluate non-target use by herbivorous biological control candidates. *BioControl*, 63(3):405–416. doi: 10.1007/s10526-018-9875-7.
- Schmidt, S., Taeger, A., Morinière, J., Liston, A., Blank, S. M., Kramp, K., Kraus, M., Schmidt, O., Heibo, E., Prous, M., Nyman, T., Malm, T., and Stahlhut, J., 2017. Identification of sawflies and horntails (Hymenoptera, Symphyta) through DNA barcodes: successes and caveats. *Molecular Ecology Resources*, 17(4):670–685. doi: 10.1111/1755-0998.12614.
- Schwarzländer, M., Hinz, H. L., Winston, R., and Day, M., 2018. Biological control of weeds: an analysis of introductions, rates of establishment and estimates of success, worldwide. *BioControl*, 63(3):319–331.
- Secord, D. and Kareiva, P., 1996. Perils and pitfalls in the host specificity paradigm. *BioScience*, 46(6):448–453.
- Shackleton, C. M., McGarry, D., Fourie, S., Gambiza, J., Shackleton, S. E., and Fabricius, C., 2007. Assessing the effects of invasive alien species on rural livelihoods: case examples and a framework from South Africa. *Human Ecology*, 35(1):113–127.
- Shackleton, S., Kirby, D., and Gambiza, J., 2011. Invasive plants—friends or foes? Contribution of prickly pear (*Opuntia ficus-indica*) to livelihoods in Makana Municipality, Eastern Cape, South Africa. *Development Southern Africa*, 28(2):177–193.
- Sheehan, M. and Potter, S., 2017. *Managing Opuntoid Cacti in Australia: Best practice control manual for Austrocyllindropuntia, Cyllindropuntia and Opuntia species*. Department of Primary Industries and Regional Development, WA, Perth. ISBN 978 0 9923083 6 0.
- Sheppard, A., Hill, R., DeClerck-Floate, R., McClay, A., Olckers, T., Quimby Jr, P., Zimmermann, H., et al., 2003. A global review of risk-benefit-cost analysis for the introduction of classical biological control agents against weeds: a crisis in the making? *Biocontrol News and Information*, 24(4).

- Silva, D., Mergulhão, C., Medeiros, L., Figueiredo, M., and Burity, H., 2013. Genetic variability of *Dactylopius opuntiae* (Hemiptera, Dactylopiidae) on forage cactus in northeast Brazil. *Genetics and Molecular Research*, 12(4):5236–5246.
- Simmons, M. P., Zhang, L.-B., Webb, C. T., and Müller, K., 2007. A penalty of using anonymous dominant markers (AFLPs, ISSRs, and RAPDs) for phylogenetic inference. *Molecular Phylogenetics and Evolution*, 42(2):528–542.
- Simpson, G. G., 1951. The species concept. *Evolution*, 5(4):285–298.
- Skibinski, D. O., Gallagher, C., and Beynon, C. M., 1994. Sex-Limited Mitochondrial DNA Transmission in the Marine Mussel *Mytilus edulis*. *Genetics*, 138(3):801–809.
- Soetaert, K. Plot3d: plotting multi-dimensional data. <https://cran.r-project.org/package=plot3D>, 2017.
- Song, H., Buhay, J. E., Whiting, M. F., and Crandall, K. A., 2008. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proceedings of the National Academy of Sciences*, 105(36):13486–13491. doi: 10.1073/pnas.0803076105.
- Sota, T., Ishikawa, R., Ujiie, M., Kusumoto, F., and Vogler, A. P., 2001. Extensive trans-species mitochondrial polymorphisms in the carabid beetles *Carabus* subgenus *Ohomopterus* caused by repeated introgressive hybridization. *Molecular Ecology*, 10(12):2833–2847. doi: 10.1046/j.0962-1083.2001.01404.x.
- Southwood, T., 1961. The number of species of insect associated with various trees. *The Journal of Animal Ecology*, pages 1–8.
- Sperling, F. A. H. and Harrison, R. G., 1994. Mitochondrial DNA variation within and between species of the *Papilio Machaon* group of Swallowtail butterflies. *Evolution*, 48(2):408–422. doi: 10.1111/j.1558-5646.1994.tb01320.x.
- Spitsyn, V. M., Kondakov, A. V., Bolotov, N. I., Thi Pham, N., Gofarov, M. Y., and Bolotov, I. N., 2018. DNA barcoding unravels contrasting evolutionary history of two widespread Asian tiger moth species during the Late Pleistocene. *Public Library of Science One*, 13(4):e0194200. doi: 10.1371/journal.pone.0194200.
- Spodek, M., Ben-Dov, Y., Protasov, A., Carvalho, C. J., and Mendel, Z., 2014. First record of *Dactylopius opuntiae* (Cockerell) (Hemiptera: Coccoidea: Dactylopiidae) from Israel. *Phytoparasitica*, 42(3):377–379.
- Stevenson, A., 1970. Strains of the grape phylloxera in Ontario with different effects on the foliage of certain grape cultivars. *Journal of Economic Entomology*, 63(1):135–138.
- Stewart, J. B., Freyer, C., Elson, J. L., and Larsson, N.-G., 2008. Purifying Selection of mtDNA and Its Implications for Understanding Evolution and Mitochondrial Disease. *Nature Reviews. Genetics*, 9(9):657–662. doi: 10.1038/nrg2396.
- Stouthamer, R., Luck, R. F., and Hamilton, W. D., 1990. Antibiotics Cause Parthenogenetic *Trichogramma* (Hymenoptera/Trichogrammatidae) to Revert to Sex. *Proceedings of the National Academy of Sciences of the United States of America*, 87(7):2424–2427.
- Suckling, D. M. and Sforza, R. F. H., 2014. What magnitude are observed non-target impacts from weed biocontrol? *Public Library of Science One*, 9(1):e84847.

- Sutton, G. F., 2019. Searching for a needle in a haystack: Where to survey for climatically-matched biological control agents for two grasses (*Sporobolus* spp.) invading Australia. *Biological Control*, 129:37–44.
- Sutton, G. F., Paterson, I. D., and 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. *Biological Control*, 114:167–175. doi: 10.1016/j.biocontrol.2017.08.015.
- Suzuki, R. and Shimodaira, H., 2006. Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*, 22(12):1540–1542.
- Taylor, P. D., 1981. Intra-Sex and Inter-Sex Sibling Interactions as Sex Ratio Determinants. *Nature*, 291(5810):64–66. doi: 10.1038/291064a0.
- Taylor, S. J., Downie, D. A., and Paterson, I. D., 2011. Genetic Diversity of Introduced Populations of the Water Hyacinth Biological Control Agent *Eccritotarsus catarinensis* (Hemiptera: Miridae). *Biological Control*, 58(3):330–336. doi: 10.1016/j.biocontrol.2011.05.008.
- Tenover, F. C., Arbeit, R. D., Goering, R. V., Mickelsen, P. A., Murray, B. E., Persing, D. H., and Swaminathan, B., 1995. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. *Journal of Clinical Microbiology*, 33(9):2233.
- Thompson, J. D., 1991. The Biology of an Invasive Plant. *BioScience*, 41(6):393–401. doi: 10.2307/1311746.
- Thorpe, W., 1930. Biological races in insects and allied groups. *Biological Reviews*, 5(3):177–212.
- Torres, J. B. and Giorgi, J. A., 2018. Management of the false carmine cochineal *Dactylopius opuntiae* (Cockerell): perspective from Pernambuco state, Brazil. *Phytoparasitica*, 46:331–340.
- Tureson, G., 1922. The genotypical response of the plant species to the habitat. *Hereditas*, 3(3):211–350.
- Turrill, W., 1946. The ecotype concept: A consideration with appreciation and criticism, especially of recent trends. *New Phytologist*, 45(1):34–43.
- Valentini, A., Pompanon, F., and Taberlet, P., 2008. DNA Barcoding for Ecologists. *Trends in Ecology & Evolution*, 24(2):110–117. doi: 10.1016/j.tree.2008.09.011.
- Valière, N., Fumagalli, L., Gielly, L., Miquel, C., Lequette, B., Poulle, M.-L., Weber, J.-M., Arlettaz, R., and Taberlet, P., 2003. Long-Distance Wolf Recolonization of France and Switzerland Inferred from Non-Invasive Genetic Sampling over a Period of 10 Years. *Animal Conservation*, 6(1):83–92. doi: 10.1017/S1367943003003111.
- Van Dam, A. R. and May, B., 2012. A new species of *Dactylopius* Costa (*Dactylopius gracilipilus* sp. nov.) (Hemiptera: Coccoidea: Dactylopiidae) from the Chihuahuan Desert, Texas, U.S.A. *Zootaxa*, 3573:33–39.
- Van Dam, A. R., Portillo Martinez, L., Chavez, A. J., and May, B. P., 2015. Range Wide Phylogeography of *Dactylopius coccus* (Hemiptera: Dactylopiidae). *Annals of the Entomological Society of America*, 108(3):299–310. doi: 10.1093/aesa/sav017.
- Van Dam, N., De Jong, T., Iwasa, Y., and Kubo, T., 1996. Optimal distribution of defences: are plants smart investors? *Functional Ecology*, pages 128–136.

- van den Berg, E. C., 2010. *Detection, quantification and monitoring Prosopis spp. in the Northern Cape Province of South Africa using Remote Sensing and GIS*. PhD thesis, North-West University, Potchefstroom.
- van Driesche, R., Hoddle, M., and Center, T., 2009. *Control of Pests and Weeds by Natural Enemies: An Introduction to Biological Control*. John Wiley & Sons. ISBN 978-1-4443-0041-3.
- Van Klinken, R. D. and Raghun, S., 2006. A scientific approach to agent selection. *Australian Journal of Entomology*, 45(4):253–258.
- Van Klinken, R. D., 1999. Host specificity testing: why do we do it and how we can do it better. In *Proceedings, host specificity testing of exotic arthropod biological control agents: the biological basis for improvement in safety, X International Symposium on Biological Control of Weeds*, pages 54–68, 1999.
- van Wilgen, B. W. and Lange, W. J. D., 2004. The Costs and Benefits of Biological Control of Invasive Alien Plants in South Africa. *African Entomology*, 19(2):504–514. doi: 10.4001/003.019.0228.
- van Wilgen, B., Richardson, D., Le Maitre, D., Marais, C., and Magadlala, D., 2001. The economic consequences of alien plant invasions: examples of impacts and approaches to sustainable management in South Africa. *Environment, Development and Sustainability*, 3(2):145–168.
- Venables, W. N. and Ripley, B. D., 2002. *Modern Applied Statistics with S*. Springer, New York, fourth edition. ISBN 0-387-95457-0.
- Vilà, M. and Hulme, P. E., 2017. Non-Native Species, Ecosystem Services, and Human Well-Being. In *Impact of Biological Invasions on Ecosystem Services*, Invading Nature - Springer Series in Invasion Ecology, pages 1–14. Springer, Cham. ISBN 978-3-319-45119-0 978-3-319-45121-3.
- Villamagna, A. M. and Murphy, B. R., 2010. Ecological and socio-economic impacts of invasive water hyacinth (*Eichhornia crassipes*): a review. *Freshwater Biology*, 55(2):282–298. doi: 10.1111/j.1365-2427.2009.02294.x.
- Volchansky, C. R., Hoffmann, J. H., and 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. *Journal of Applied Ecology*, 36(1):85–91. doi: 10.1046/j.1365-2664.1999.00381.x.
- Wallace, R. S. and Gibson, A. C., 2002. Evolution and systematics. In Nobel, P. S., editor, *Cacti: Biology and Uses*, pages 1–22. California University Press.
- Walters, M., Figueiredo, E., Crouch, N., Winter, P., Smith, G., Zimmermann, H., and Mashope, B., 2011. Naturalised and invasive succulents of Southern Africa. *ABC Taxa*, 11:91–217.
- Wan, Q.-H., Wu, H., Fujihara, T., and Fang, S.-G., 2004. Which genetic marker for which conservation genetics issue? *Electrophoresis*, 25(14):2165–2176.
- Wang, J., 2017. The computer program STRUCTURE for assigning individuals to populations: easy to use but easier to misuse. *Molecular Ecology Resources*, 17(5):981–990.
- Wapshere, A., 1974. A strategy for evaluating the safety of organisms for biological weed control. *Annals of Applied Biology*, 77(2):201–211.
- Wapshere, A., 1985. Effectiveness of biological control agents for weeds: present quandaries. *Agriculture, Ecosystems & Environment*, 13(3-4):261–280.

- Waugh, J., 2007. DNA Barcoding in Animal Species: Progress, Potential and Pitfalls. *BioEssays*, 29 (2):188–197. doi: 10.1002/bies.20529.
- Weir, J. T. and Schluter, D., 2008. Calibrating the Avian Molecular Clock. *Molecular Ecology*, 17(10): 2321–2328. doi: 10.1111/j.1365-294X.2008.03742.x.
- Wilcox, T. P., Zwickl, D. J., Heath, T. A., and Hillis, D. M., 2002. Phylogenetic relationships of the dwarf boas and a comparison of bayesian and bootstrap measures of phylogenetic support. *Molecular Phylogenetics and Evolution*, 25(2):361–371.
- Willerslev, E., Hansen, A. J., Binladen, J., Brand, T. B., Gilbert, M. T. P., Shapiro, B., Bunce, M., Wiuf, C., Gilichinsky, D. A., and Cooper, A., 2003. Diverse Plant and Animal Genetic Records from Holocene and Pleistocene Sediments. *Science*, 300(5620):791–795. doi: 10.1126/science.1084114.
- Willette, D. A., Simmonds, S. E., Cheng, S. H., Esteves, S., Kane, T. L., Nuetzel, H., Pilaud, N., Rachmawati, R., and Barber, P. H., 2017. Using DNA Barcoding to Track Seafood Mislabeling in Los Angeles Restaurants. *Conservation Biology: The Journal of the Society for Conservation Biology*, 31(5):1076–1085. doi: 10.1111/cobi.12888.
- Williams, S. T. and Knowlton, N., 2001. Mitochondrial Pseudogenes Are Pervasive and Often Insidious in the Snapping Shrimp Genus *Alpheus*. *Molecular Biology and Evolution*, 18(8):1484–1493. doi: 10.1093/oxfordjournals.molbev.a003934.
- Wilm, A., Higgins, D. G., and Notredame, C., 2008. R-Coffee: a method for multiple alignment of non-coding RNA. *Nucleic Acids Research*, 36(9):e52–e52.
- Wilson, F., 1961. A review of the biological control of insects and weeds in Australia and Australian New Guinea. *Entomophaga*, 6(1):76. doi: <https://doi.org/10.1007/BF02373207>.
- Wink, M., 2006. Use of DNA markers to study bird migration. *Journal of Ornithology*, 147(2):234–244.
- Winston, R. L., Schwarzländer, M., Hinz, H. L., Day, M. D., Cock, M. J. W., and Julien, M. H., 2014. Biological control of weeds: a world catalogue of agents and their target weeds. *Biological control of weeds: a world catalogue of agents and their target weeds.*, (5).
- Wolfe, A. D., 2005. ISSR techniques for evolutionary biology. In *Methods in Enzymology*, volume 395, pages 134–144. Elsevier.
- Wolfe, A. D., Xiang, Q.-Y., and Kephart, S. R., 1998. Assessing Hybridization in Natural Populations of *Penstemon* (Scrophulariaceae) Using Hypervariable Intersimple Sequence Repeat (ISSR) Bands. *Molecular Ecology*, 7(9):1107–1125. doi: 10.1046/j.1365-294x.1998.00425.x.
- Wolff, J. N., Nafisinia, M., Sutovsky, P., and Ballard, J. W. O., 2013. Paternal Transmission of Mitochondrial DNA as an Integral Part of Mitochondrial Inheritance in Metapopulations of *Drosophila simulans*. *Heredity*, 110(1):57–62. doi: 10.1038/hdy.2012.60.
- Wrench, D. L., Kethley, J. B., and Norton, R. A., 1994. Cytogenetics of Holokinetic Chromosomes and Inverted Meiosis: Keys to the Evolutionary Success of Mites, with Generalizations on Eukaryotes. In *Mites*, pages 282–343. Springer, Boston, MA. ISBN 978-1-4613-6012-4 978-1-4615-2389-5.
- Wright, E. S., 2016. Using DECIPHER v2. 0 to analyze big biological sequence data in R. *R Journal*, 8(1).
- Xia, X., 2018. DAMBE7: New and improved tools for data analysis in molecular biology and evolution. *Molecular Biology and Evolution*, 35(6):1550–1552.

- Xia, X. and Lemey, P., 2009. Assessing substitution saturation with DAMBE. In Lemey, P., Salemi, M., and Vandamme, A. M., editors, *The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny*, pages 615–630. Cambridge University Press.
- Xia, X., Xie, Z., Salemi, M., Chen, L., and Wang, Y., 2003. An index of substitution saturation and its application. *Molecular Phylogenetics and Evolution*, 26(1):1–7.
- Xiong, H., Barker, S. C., Burger, T. D., Raoult, D., and Shao, R., 2013. Heteroplasmy in the Mitochondrial Genomes of Human Lice and Ticks Revealed by High Throughput Sequencing. *Public Library of Science One*, 8(9):e73329. doi: 10.1371/journal.pone.0073329.
- Xu, S.-Z., Li, Z.-Y., and Jin, X.-H., 2017. DNA Barcoding of Invasive Plants in China: A Resource for Identifying Invasive Plants. *Molecular Ecology Resources*, 18(1):128–136. doi: 10.1111/1755-0998.12715.
- Zachariades, C. Biological control of invasive alien plants in South Africa: a list of all insects, mites and pathogens considered as biological control agents from 1913-2018, 2018. URL <http://www.arc.agric.za/arc-ppri/Documents/Table1-NaturalEnemiesAll.pdf>. Last accessed 09 May 2019.
- Zachariades, C., Paterson, I. D., Strathie, L. W., Hill, M. P., and van Wilgen, B. W., 2017. Assessing the status of biological control as a management tool for suppression of invasive alien plants in South Africa. *Bothalia-African Biodiversity & Conservation*, 47(2):1–19.
- Zhang, D. X. and Hewitt, G. M., 1996. Nuclear integrations: challenges for mitochondrial DNA markers. *Trends in Ecology & Evolution*, 11(6):247–51.
- Zhao, Y., Wang, Y., Wang, L., and Zhang, D., 2019. Molecular identification of mung bean accessions (*Vigna radiata* L.) from Northeast China using capillary electrophoresis with fluorescence-labeled SSR markers. *Food and Energy Security*, page e182.
- Zietkiewicz, E., Rafalski, A., and Labuda, D., 1994. Genome Fingerprinting by Simple Sequence Repeat (SSR)-Anchored Polymerase Chain Reaction Amplification. *Genomics*, 20(2):176–183. doi: 10.1006/geno.1994.1151.
- Zimmermann, H. G. and Granata, G., 2002. Insect pests and diseases. In Nobel, P. S., editor, *Cacti: Biology and Uses*, pages 231–254. California University Press, Berkeley, CA.
- Zimmermann, H. G. and Moran, V. C., 1991. Biological Control of Prickly Pear, *Opuntia ficus-indica* (Cactaceae), in South Africa. *Agriculture, Ecosystems & Environment*, 37(1):29–35. doi: 10.1016/0167-8809(91)90137-M.
- Zimmermann, H. G., Moran, V. C., and Hoffmann, J. H., 2004. Biological Control in the Management of Invasive Alien Plants in South Africa, and the Role of the Working for Water Programme : Working for Water. *South African Journal of Science*, 100(1-2):34–40.
- Zimmermann, H. G., Moran, V. C., and Hoffmann, J. H., 2009. Invasive cactus species (Cactaceae). In Muniappan, R., Reddy, G. V. P., and Raman, A., editors, *Biological Control of Tropical Weeds using Arthropods*, pages 108–129. Cambridge University Press, Cambridge, United Kingdom.
- Zimmermann, H. and van de Venter, H., 1981. Jointed cactus. Technical report, Sub-division Weeds Research: Plant Protection Research Institute.

- Zouros, E., Freeman, K. R., Ball, A. O., and Pogson, G. H., 1992. Direct evidence for extensive paternal mitochondrial DNA inheritance in the marine mussel *Mytilus*. *Nature*, 359(6394):412–414. doi: 10.1038/359412a0.
- Zouros, E., Oberhauser Ball, A., Saavedra, C., and Freeman, K. R., 1994. An Unusual Type of Mitochondrial DNA Inheritance in the Blue Mussel *Mytilus*. *Proceedings of the National Academy of Sciences of the United States of America*, 91(16):7463–7467.
- 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.

# 5

## Appendix A

### 5.1 Extraction protocol

1. 180  $\mu\text{L}$  buffer ATL and 15  $\mu\text{L}$  proteinase K was added to tissue (after all the ethanol had evaporated off), and incubated in a heatblock at 56°C overnight.
2. Digested tissue was centrifuged for 5 min at 16000 RCF (relative centrifugal force).
3. The supernatant was pipetted into a clean tube, and 65  $\mu\text{L}$  NaCl (5M) added. The sample was vortexed for 30s, and then centrifuged for 5 min.
4. The supernatant was pipetted into another clean tube (leaving behind the white salt residue), and 150  $\mu\text{L}$  isopropanol (98%) added. The sample was gently mixed, and left in a freezer for at least 2 hr.
5. The sample was centrifuged for 5 min, and the liquid removed (leaving behind the DNA pellet).
6. 250  $\mu\text{L}$  cooled ethanol (80%) was added, followed by 30s of vortexing and 5 min centrifuging. The ethanol was then removed, and step 6 was repeated.
7. After removing the ethanol, the sample was left in a heatblock at 40°C for 15 min in order to dry the DNA pellet.
8. The sample was suspended in 100  $\mu\text{L}$  buffer AE, and stored at -18°C.

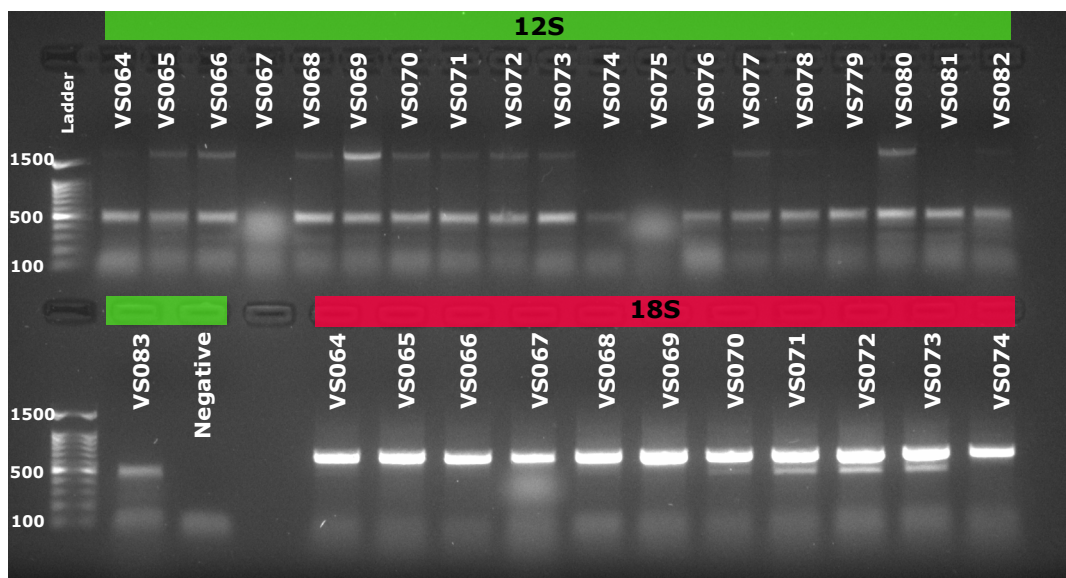
### 5.2 DNA sequences

#### 5.2.1 Submission to Genbank

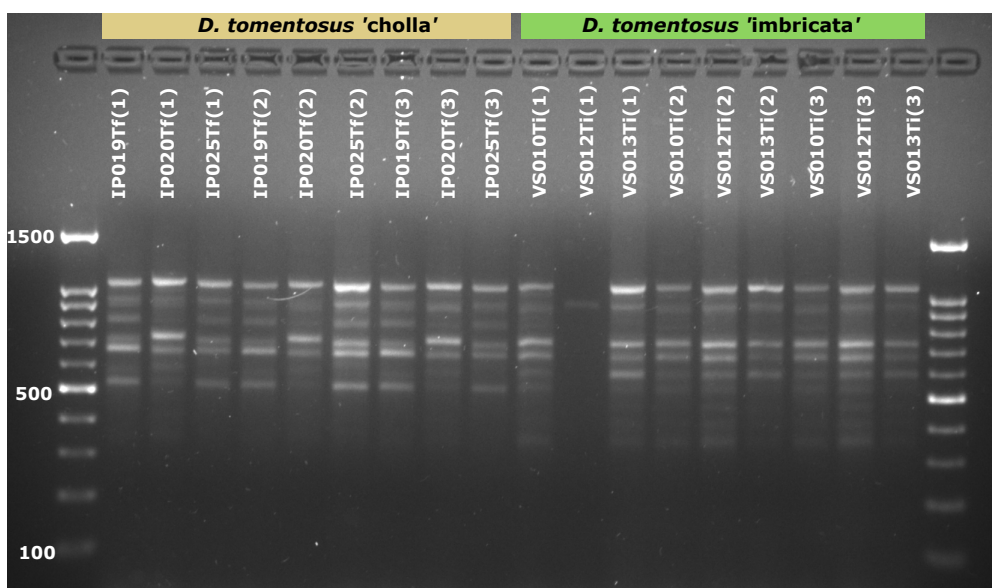
Sequences were uploaded to Genbank via the submission portal available at <https://submit.ncbi.nlm.nih.gov/>. The COI protein-coding sequences were uploaded via BankIt (accessed from the same Genbank submission portal). This required information regarding the start codon position, which was found by ob-

taining the longest open reading frame using ORFfinder (<https://www.ncbi.nlm.nih.gov/orffinder/>). The genetic code was set to ‘Invertebrate Mitochondrial’ and ORF codon start codon to ‘ATG and alternative initiation codons’.

## 5.2.2 Gel images



**Figure A1:** 12S and 18S PCR products visualised on 1% agarose gel. Note the double bands in many of the 12S products; at the 500 and 1500 base pair mark.



**Figure A2:** ISSR banding pattern on a 1.5% agarose gel for the two *D. tomentosus* lineages (‘cholla’ and ‘imbricata’), run as a preliminary test before fragment analysis. Ladders are on the far left and right, marked at 100, 500 and 1500 base pair sizes.

**Table A1:** Genbank accession numbers for 12S, 18S and COI genetic sequences; with the extraction sample ID, species, lineage, locality and host plant information. CT Bot. Gard. = Cape Town Botanical Gardens, EC = Eastern Cape, KNP = Kruger National Park, MRF = Mass Rearing Facility. Black dots indicate samples collected in the native range in 2017.

12S	18S	COI DTOMf	COI PcoF1	Extraction ID	Species	Lineage	Locality	Host plant
MN220012				VS021Au	<i>D. austrinus</i>		South Africa: Uitenhage	<i>Opuntia aurantiaca</i>
MN220013	MN310065			IP001Au	<i>D. austrinus</i>		South Africa: Uitenhage	<i>Opuntia aurantiaca</i>
MN220014	MN310073			VS001Au	<i>D. austrinus</i>		South Africa: Uitenhage	<i>Opuntia aurantiaca</i>
MN220038	MN310132			VS022Au	<i>D. austrinus</i>		South Africa: Uitenhage	<i>Opuntia aurantiaca</i>
MN220079				VS089Au	<i>D. austrinus</i>		South Africa: Uitenhage	<i>Opuntia aurantiaca</i>
MN220091	MN310131			VS090Au	<i>D. austrinus</i>		South Africa: Uitenhage	<i>Opuntia aurantiaca</i>
MN220130	MN310178			Au4	<i>D. austrinus</i>		Australia: Queensland	<i>Opuntia aurantiaca</i>
MN220131	MN310179			Au3	<i>D. austrinus</i>		Australia: Queensland	<i>Opuntia aurantiaca</i>
MN220132	MN310149			Au1	<i>D. austrinus</i>		Australia: Queensland	<i>Opuntia aurantiaca</i>
	MN310069			IP021Au	<i>D. austrinus</i>		South Africa: Uitenhage	<i>Opuntia aurantiaca</i>
	MN310130			VS091Au	<i>D. austrinus</i>		South Africa: Uitenhage	<i>Opuntia aurantiaca</i>
MN220010				VS020C	<i>D. ceylonicus</i>		South Africa: Uitenhage	<i>Opuntia monacantha</i>
MN220011	MN310074			VS002C	<i>D. ceylonicus</i>		South Africa: Uitenhage	<i>Opuntia monacantha</i>
MN220037	MN310134			VS019C	<i>D. ceylonicus</i>		South Africa: Uitenhage	<i>Opuntia monacantha</i>
MN220039	MN310133			VS021C	<i>D. ceylonicus</i>		South Africa: Uitenhage	<i>Opuntia monacantha</i>
MN220122	MN310173			cey4	<i>D. ceylonicus</i>		Australia: Queensland	<i>Opuntia monacantha</i>
MN220123	MN310152			cey3	<i>D. ceylonicus</i>		Australia: Queensland	<i>Opuntia monacantha</i>
MN220124	MN310160			cey2	<i>D. ceylonicus</i>		Australia: Queensland	<i>Opuntia monacantha</i>
	MN310070			IP022C	<i>D. ceylonicus</i>		South Africa: Uitenhage	<i>Opuntia monacantha</i>
	MN310081			VS020C	<i>D. ceylonicus</i>		South Africa: Uitenhage	<i>Opuntia monacantha</i>
	MN310174			cey1	<i>D. ceylonicus</i>		Australia: Queensland	<i>Opuntia monacantha</i>
	MN310037		MN372226	CSW001H	<i>D. confertus</i>		Namibia, Windhoek	<i>Harrisia sp.</i>
	MN310038		MN372227	CSW002H	<i>D. confertus</i>		Namibia, Windhoek	<i>Harrisia sp.</i>
MN220015	MN310099		MN372274	VS100Con_s	<i>D. confusus</i>		USA: Arizona, Tucson ●	<i>Opuntia santa-rita</i>
MN220016	MN310100			VS101Con_e	<i>D. confusus</i>		USA: Arizona, Tucson	<i>Opuntia engelmannii</i>
MN220017	MN310101			VS102Con_e	<i>D. confusus</i>		USA: Arizona, Tucson	<i>Opuntia engelmannii</i>
MN220018	MN310102			VS103Con_e	<i>D. confusus</i>		USA: Arizona, Tucson	<i>Opuntia engelmannii</i>
MN220019	MN310103			VS104Con_e	<i>D. confusus</i>		USA: Arizona, Tucson	<i>Opuntia engelmannii</i>
MN220029			MN372263	VS035O.4PM	<i>D. confusus</i>		USA: Arizona, Four Peaks Mt. ●	<i>Opuntia engelmannii</i>
MN220030			MN372264	VS036O.4PM	<i>D. confusus</i>		USA: Arizona, Four Peaks Mt. ●	<i>Opuntia engelmannii</i>
MN220031	MN310091		MN372265	VS037O.4PM	<i>D. confusus</i>		USA: Arizona, Four Peaks Mt. ●	<i>Opuntia engelmannii</i>
MN220032	MN310092		MN372266	VS038O.lc	<i>D. confusus</i>		USA: New Mexico, Las Cruces ●	<i>Opuntia engelmannii</i>
MN220033	MN310093		MN372267	VS039O.lc	<i>D. confusus</i>		USA: New Mexico, Las Cruces ●	<i>Opuntia engelmannii</i>
MN220034	MN310094		MN372268	VS040O.lc	<i>D. confusus</i>		USA: New Mexico, Las Cruces ●	<i>Opuntia engelmannii</i>
MN220054	MN310054		MN372240	VS072u	<i>D. confusus</i>		USA: Texas, Laredo ●	<i>Opuntia engelmannii</i>
MN220055	MN310055		MN372241	VS073u	<i>D. confusus</i>		USA: Texas, Laredo ●	<i>Opuntia engelmannii</i>
MN220056	MN310056		MN372242	VS074u	<i>D. confusus</i>		USA: Texas, Laredo ●	<i>Opuntia engelmannii</i>
MN220059	MN310060		MN372246	VS078u	<i>D. confusus</i>		USA: Arizona, Tucson ●	<i>Opuntia engelmannii</i>
MN220060	MN310063		MN372249	VS081u	<i>D. confusus</i>		USA: New Mexico, Las Cruces ●	<i>Opuntia engelmannii</i>
MN220063	MN310061		MN372247	VS079u	<i>D. confusus</i>		USA: Arizona, Tucson ●	<i>Opuntia engelmannii</i>
MN220064	MN310062		MN372248	VS080u	<i>D. confusus</i>		USA: New Mexico, Las Cruces ●	<i>Opuntia engelmannii</i>
	MN310057		MN372243	VS075u	<i>D. confusus</i>		USA: Texas, Laredo ●	<i>Opuntia engelmannii</i>
MN219994	MN310078			VS015Ofi	<i>D. opuntiae</i>	ficus	South Africa: Uitenhage	<i>Opuntia ficus-indica</i>
MN219995	MN310079		MN372255	VS016Ofi	<i>D. opuntiae</i>	ficus	South Africa: Uitenhage	<i>Opuntia ficus-indica</i>
MN219996	MN310080		MN372256	VS017Ofi	<i>D. opuntiae</i>	ficus	South Africa: Uitenhage	<i>Opuntia ficus-indica</i>
MN219997	MN310095		MN372269	VS044Ofi	<i>D. opuntiae</i>	ficus	Namibia: Auros farm	<i>Opuntia ficus-indica</i>
MN219998	MN310096		MN372271	VS046Ofi	<i>D. opuntiae</i>	ficus	Namibia: Auros farm	<i>Opuntia ficus-indica</i>
MN219999	MN310142		MN372270	VS045Ofi	<i>D. opuntiae</i>	ficus	Namibia: Auros farm	<i>Opuntia ficus-indica</i>
MN220005	MN310066			IP003Os	<i>D. opuntiae</i>		South Africa: Uitenhage, MRF	<i>Opuntia ficus-indica</i>
MN220006	MN310098		MN372273	VS054Os	<i>D. opuntiae</i>	ficus	Namibia: Auros farm	<i>Opuntia stricta</i>
MN220007	MN310139			VS055Os	<i>D. opuntiae</i>	ficus	Namibia: Auros farm	<i>Opuntia stricta</i>
MN220008	MN310097		MN372272	VS053Os	<i>D. opuntiae</i>	ficus	Namibia: Auros farm	<i>Opuntia stricta</i>
MN220009	MN310075		MN372252	VS004Os	<i>D. opuntiae</i>		South Africa: Uitenhage, MRF	<i>Opuntia ficus-indica</i>
MN220020	MN310082		MN372257	VS026O.flag	<i>D. opuntiae</i>		USA: Arizona, Flagstaff ●	<i>Opuntia engelmannii</i>

MN220021	MN310083	MN372258	VS027O.flag	<i>D. opuntiae</i>	USA: Arizona, Flagstaff ●	<i>Opuntia engelmannii</i>
MN220022	MN310084		VS028O.flag	<i>D. opuntiae</i>	USA: Arizona, Flagstaff ●	<i>Opuntia engelmannii</i>
MN220023	MN310085		VS029O.tuc	<i>D. opuntiae</i>	USA: Arizona, Tucson ●	<i>Opuntia engelmannii</i>
MN220024	MN310086	MN372259	VS030O.tuc	<i>D. opuntiae</i>	USA: Arizona, Tucson ●	<i>Opuntia engelmannii</i>
MN220025	MN310087	MN372260	VS031O.tuc	<i>D. opuntiae</i>	USA: Arizona, Tucson ●	<i>Opuntia engelmannii</i>
MN220026	MN310088	MN372261	VS032O.sed	<i>D. opuntiae</i>	USA: Arizona, Sedona ●	<i>Opuntia engelmannii</i>
MN220027	MN310089		VS033O.sed	<i>D. opuntiae</i>	USA: Arizona, Sedona ●	<i>Opuntia engelmannii</i>
MN220028	MN310090	MN372262	VS034O.sed	<i>D. opuntiae</i>	USA: Arizona, Sedona ●	<i>Opuntia engelmannii</i>
MN220035	MN310039	MN372228	CSW003E	<i>D. opuntiae</i>	Namibia, Windhoek	<i>Opuntia engelmannii</i>
MN220036	MN310040		CSW004E	<i>D. opuntiae</i>	Namibia, Windhoek	<i>Opuntia engelmannii</i>
MN220043	MN310041	MN372229	VS059Os	<i>D. opuntiae</i>	South Africa: Uitenhage, MRF	<i>Opuntia ficus-indica</i>
MN220044	MN310042	MN372230	VS060Os	<i>D. opuntiae</i>	South Africa: Uitenhage, MRF	<i>Opuntia ficus-indica</i>
MN220045	MN310043	MN372231	VS061Os	<i>D. opuntiae</i>	South Africa: Uitenhage, MRF	<i>Opuntia ficus-indica</i>
MN220046	MN310044	MN372232	VS062Os	<i>D. opuntiae</i>	South Africa: Uitenhage, MRF	<i>Opuntia ficus-indica</i>
MN220047	MN310045	MN372233	VS063Os	<i>D. opuntiae</i>	South Africa: Uitenhage, MRF	<i>Opuntia ficus-indica</i>
MN220048	MN310046	MN372234	VS064u	<i>D. opuntiae</i>	USA: Texas, Moor Airbase ●	<i>Opuntia engelmannii</i>
MN220049	MN310047	MN372235	VS065u	<i>D. opuntiae</i>	USA: Texas, Moor Airbase ●	<i>Opuntia engelmannii</i>
MN220050	MN310048	MN372236	VS066u	<i>D. opuntiae</i>	USA: Texas, Port Isabel ●	<i>Opuntia engelmannii</i>
MN220051	MN310050	MN372238	VS068u	<i>D. opuntiae</i>	USA: Texas, Laredo ●	<i>Opuntia engelmannii</i>
MN220052	MN310051		VS069u	<i>D. opuntiae</i>	USA: Texas, Laredo ●	<i>Opuntia engelmannii</i>
MN220053	MN310053	MN372239	VS071u	<i>D. opuntiae</i>	USA: Texas, Webb ●	<i>Opuntia engelmannii</i>
MN220057	MN310058	MN372244	VS076u	<i>D. opuntiae</i>	USA: California, Rincon ●	<i>Opuntia engelmannii</i>
MN220058	MN310059	MN372245	VS077u	<i>D. opuntiae</i>	USA: California, Rincon ●	<i>Opuntia engelmannii</i>
MN220061			VS083u	<i>D. opuntiae</i>	USA: Texas, Pecos County ●	<i>Opuntia engelmannii</i>
MN220062	MN310052		VS070u	<i>D. opuntiae</i>	USA: Texas, Webb ●	<i>Opuntia engelmannii</i>
MN220065	MN310104		VS086uOE	<i>D. opuntiae</i>	South Africa: Cookhouse, EC	<i>Opuntia engelmannii</i>
MN220066	MN310116		VS087uOE	<i>D. opuntiae</i>	South Africa: Cookhouse, EC	<i>Opuntia engelmannii</i>
MN220067	MN310115		VS088uOE	<i>D. opuntiae</i>	South Africa: Cookhouse, EC	<i>Opuntia engelmannii</i>
MN220074	MN310107		VS106V	<i>D. opuntiae</i>	South Africa: EC	<i>Opuntia engelmannii</i>
MN220075	MN310106		VS107V	<i>D. opuntiae</i>	South Africa: EC	<i>Opuntia engelmannii</i>
MN220076	MN310105		VS108V	<i>D. opuntiae</i>	South Africa: EC	<i>Opuntia engelmannii</i>
MN220077	MN310125		P2S4	<i>D. opuntiae</i>	South Africa: KNP	<i>Opuntia stricta</i>
MN220078	MN310122	MN372278	P3S2	<i>D. opuntiae</i>	South Africa: KNP	<i>Opuntia stricta</i>
MN220080	MN310145		BCCS1	<i>D. opuntiae</i>	South Africa: KNP, Skukuza	<i>Opuntia stricta</i>
MN220081	MN310127	MN372275	P1S4	<i>D. opuntiae</i>	South Africa: KNP	<i>Opuntia stricta</i>
MN220082			VS054Os	<i>D. opuntiae</i>	South Africa: Namibia	<i>Opuntia stricta</i>
MN220083	MN310123		P3S1	<i>D. opuntiae</i>	South Africa: KNP	<i>Opuntia stricta</i>
MN220084	MN310128	MN372279	P1S3	<i>D. opuntiae</i>	South Africa: KNP	<i>Opuntia stricta</i>
MN220085	MN310119	MN372277	VS151H	<i>D. opuntiae</i>	South Africa	<i>Opuntia stricta</i>
MN220086	MN310143	MN372280	BCCS4	<i>D. opuntiae</i>	South Africa: KNP, Skukuza	<i>Opuntia stricta</i>
MN220087	MN310124		P2S5	<i>D. opuntiae</i>	South Africa: KNP	<i>Opuntia stricta</i>
MN220088	MN310129		P1S1	<i>D. opuntiae</i>	South Africa: KNP	<i>Opuntia stricta</i>
MN220089	MN310121		VS149K	<i>D. opuntiae</i>	South Africa	<i>Opuntia stricta</i>
MN220090	MN310144		BCCS2	<i>D. opuntiae</i>	South Africa: KNP, Skukuza	<i>Opuntia stricta</i>
MN220092			VS146K	<i>D. opuntiae</i>	South Africa	<i>Opuntia stricta</i>
MN220093	MN310126		P2S2	<i>D. opuntiae</i>	South Africa: KNP	<i>Opuntia stricta</i>
MN220094	MN310118	MN372276	VS152H	<i>D. opuntiae</i>	South Africa	<i>Opuntia stricta</i>
MN220095	MN310117		VS154H	<i>D. opuntiae</i>	South Africa	<i>Opuntia stricta</i>
MN220096			Ah2	<i>D. opuntiae</i>	Saudi Arabia: Jazan, Alaredah	<i>Opuntia stricta</i>
MN220097	MN310163		op4	<i>D. opuntiae</i>	Australia: Queensland	<i>Opuntia stricta</i>
MN220098	MN310147		op3	<i>D. opuntiae</i>	Australia: Queensland	<i>Opuntia stricta</i>
MN220099	MN310162		op1	<i>D. opuntiae</i>	Australia: Queensland	<i>Opuntia stricta</i>
MN220100			JC3	<i>D. opuntiae</i>	Saudi Arabia: Jazan, Jazan City	<i>Opuntia stricta</i>
MN220101			JC2	<i>D. opuntiae</i>	Saudi Arabia: Jazan, Jazan City	<i>Opuntia stricta</i>
MN220106			H4	<i>D. opuntiae</i>	Saudi Arabia: Jazan, Hurob	<i>Opuntia stricta</i>
MN220107			H1	<i>D. opuntiae</i>	Saudi Arabia: Jazan, Hurob	<i>Opuntia stricta</i>
MN220108	MN310155		eng1	<i>D. opuntiae</i>	Australia: Queensland	<i>Opuntia engelmannii</i>
MN220111	MN310166		eng4	<i>D. opuntiae</i>	Australia: Queensland	<i>Opuntia engelmannii</i>
MN220133			AL3	<i>D. opuntiae</i>	Saudi Arabia: Jazan, Aldyer	<i>Opuntia stricta</i>
MN220134			AL2	<i>D. opuntiae</i>	Saudi Arabia: Jazan, Aldyer	<i>Opuntia stricta</i>
MN220135			Ah4	<i>D. opuntiae</i>	Saudi Arabia: Jazan, Alaredah	<i>Opuntia stricta</i>
	MN310049	MN372237	VS067u	<i>D. opuntiae</i>	USA: Texas, Port Isabel	<i>Opuntia engelmannii</i>

	MN310064		MN372250	VS082u	<i>D. opuntiae</i>	USA: Texas, Pecos County	<i>Opuntia engelmannii</i>
	MN310071			IP024Os	<i>D. opuntiae</i>	South Africa: Uitenhage, MRF	<i>Opuntia ficus-indica</i>
	MN310120			VS150K	<i>D. opuntiae</i>	South Africa	<i>Opuntia stricta</i>
	MN310140			CVS262	<i>D. opuntiae</i>	South Africa: Uitenhage	<i>Opuntia stricta</i>
	MN310141			CVS261	<i>D. opuntiae</i>	South Africa: Uitenhage	<i>Opuntia stricta</i>
	MN310167			eng3	<i>D. opuntiae</i>	Australia: Queensland	<i>Opuntia engelmannii</i>
			MN372253	VS008Ofi	<i>D. opuntiae</i>	South Africa: Uitenhage	<i>Opuntia ficus-indica</i>
MN220000	MN310137			VS014Ti	<i>D. tomentosus</i>	South Africa: Uitenhage	<i>Cylindropuntia imbricata</i>
MN220001	MN310077			VS013Ti	<i>D. tomentosus</i>	South Africa: Uitenhage	<i>Cylindropuntia imbricata</i>
MN220002	MN310076		MN372254	VS012Ti	<i>D. tomentosus</i>	South Africa: Uitenhage	<i>Cylindropuntia imbricata</i>
MN220102	MN310164			imb4	<i>D. tomentosus</i>	Australia: Queensland	<i>Cylindropuntia imbricata</i>
MN220103	MN310165			imb3	<i>D. tomentosus</i>	Australia: Queensland	<i>Cylindropuntia imbricata</i>
MN220104	MN310156			imb2	<i>D. tomentosus</i>	Australia: Queensland	<i>Cylindropuntia imbricata</i>
MN220105	MN310148			imb1	<i>D. tomentosus</i>	Australia: Queensland	<i>Cylindropuntia imbricata</i>
			MN372299	IP005Ti	<i>D. tomentosus</i>	South Africa: Uitenhage	<i>Cylindropuntia imbricata</i>
MN220041	MN310138			VS010Ti	<i>D. tomentosus</i>	South Africa: Uitenhage	<i>Cylindropuntia imbricata</i>
MN220042	MN310136			IP016Ti	<i>D. tomentosus</i>	South Africa: Uitenhage	<i>Cylindropuntia imbricata</i>
MN220003	MN310072		MN372251	IP025Tf	<i>D. tomentosus</i>	South Africa: CT Bot. Gard.	<i>C. fulgida</i> var. <i>mamm.</i>
MN220004	MN310135			VS005Tf	<i>D. tomentosus</i>	South Africa: CT Bot. Gard.	<i>C. fulgida</i> var. <i>mamm.</i>
MN220040	MN310068	MN372298		IP019Tf	<i>D. tomentosus</i>	South Africa: CT Bot. Gard.	<i>C. fulgida</i> var. <i>mamm.</i>
MN220068	MN310114			VS098Tf	<i>D. tomentosus</i>	South Africa: CT Bot. Gard.	<i>C. fulgida</i> var. <i>mamm.</i>
MN220069	MN310113			VS099Tf	<i>D. tomentosus</i>	South Africa: CT Bot. Gard.	<i>C. fulgida</i> var. <i>mamm.</i>
MN220109				Ekh6	<i>D. tomentosus</i>	South Africa: Jansenville, EC	<i>Cylindropuntia fulgida</i>
MN220110				EKh5	<i>D. tomentosus</i>	South Africa: Jansenville, EC	<i>Cylindropuntia fulgida</i>
MN220112				Ekh4	<i>D. tomentosus</i>	South Africa: Jansenville, EC	<i>Cylindropuntia fulgida</i>
MN220119		MN372287		chol4	<i>D. tomentosus</i>	Australia: Queensland	<i>Cylindropuntia fulgida</i>
MN220120	MN310153	MN372288		chol3	<i>D. tomentosus</i>	Australia: Queensland	<i>Cylindropuntia fulgida</i>
MN220121	MN310172	MN372290		chol1	<i>D. tomentosus</i>	Australia: Queensland	<i>Cylindropuntia fulgida</i>
	MN310161	MN372289		chol2	<i>D. tomentosus</i>	Australia: Queensland	<i>C. fulgida</i> var. <i>mamm.</i>
	MN310067			IP005Tf	<i>D. tomentosus</i>	South Africa: CT Bot. Gard.	<i>C. fulgida</i> var. <i>mamm.</i>
MN220070	MN310112			VS101Tp	<i>D. tomentosus</i>	Australia: Queensland	<i>Cylindropuntia pallida</i>
MN220071	MN310111			VS102Tp	<i>D. tomentosus</i>	Australia: Queensland	<i>Cylindropuntia pallida</i>
MN220072	MN310109			VS104Tp	<i>D. tomentosus</i>	Australia: Queensland	<i>Cylindropuntia pallida</i>
MN220073	MN310108			VS105Tp	<i>D. tomentosus</i>	Australia: Queensland	<i>Cylindropuntia pallida</i>
MN220125	MN310175	MN372291		Ca4	<i>D. tomentosus</i>	Australia: Queensland	<i>Cylindropuntia pallida</i>
MN220126	MN310151	MN372292		Ca3	<i>D. tomentosus</i>	Australia: Queensland	<i>Cylindropuntia pallida</i>
MN220127	MN310158	MN372293		Ca2	<i>D. tomentosus</i>	Australia: Queensland	<i>Cylindropuntia pallida</i>
MN220128	MN310176	MN372294		Ca1	<i>D. tomentosus</i>	Australia: Queensland	<i>Cylindropuntia pallida</i>
	MN310110			VS103Tp	<i>D. tomentosus</i>	Australia: Queensland	<i>Cylindropuntia pallida</i>
MN220113	MN310146	MN372281		ech4	<i>D. tomentosus</i>	Australia: Queensland	<i>C. ech. x C. acanth.</i>
MN220114	MN310168	MN372282		ech3	<i>D. tomentosus</i>	Australia: Queensland	<i>C. ech. x C. acanth.</i>
MN220115		MN372283		ech2	<i>D. tomentosus</i>	Australia: Queensland	<i>C. ech. x C. acanth.</i>
MN220118	MN310169	MN372284		ech1	<i>D. tomentosus</i>	Australia: Queensland	<i>C. ech. x C. acanth.</i>
MN220116	MN310157	MN372285		cy13	<i>D. tomentosus</i>	Australia: Queensland	<i>Cylindropuntia sp.</i>
MN220117	MN310170	MN372286		cy12	<i>D. tomentosus</i>	Australia: Queensland	<i>Cylindropuntia sp.</i>
	MN310154			cy14	<i>D. tomentosus</i>	Australia: Queensland	<i>Cylindropuntia sp.</i>
	MN310171			cy11	<i>D. tomentosus</i>	Australia: Queensland	<i>Cylindropuntia sp.</i>
MN220129	MN310159	MN372297		Big2	<i>D. tomentosus</i>	Australia: Queensland	<i>Cylindropuntia bigelovii</i>
	MN310150	MN372296		Big3	<i>D. tomentosus</i>	Australia: Queensland	<i>Cylindropuntia bigelovii</i>
	MN310177	MN372295		Big5	<i>D. tomentosus</i>	Australia: Queensland	<i>Cylindropuntia bigelovii</i>

**Table A2:** Sample IDs for additional DNA extractions used for ISSR analyses.

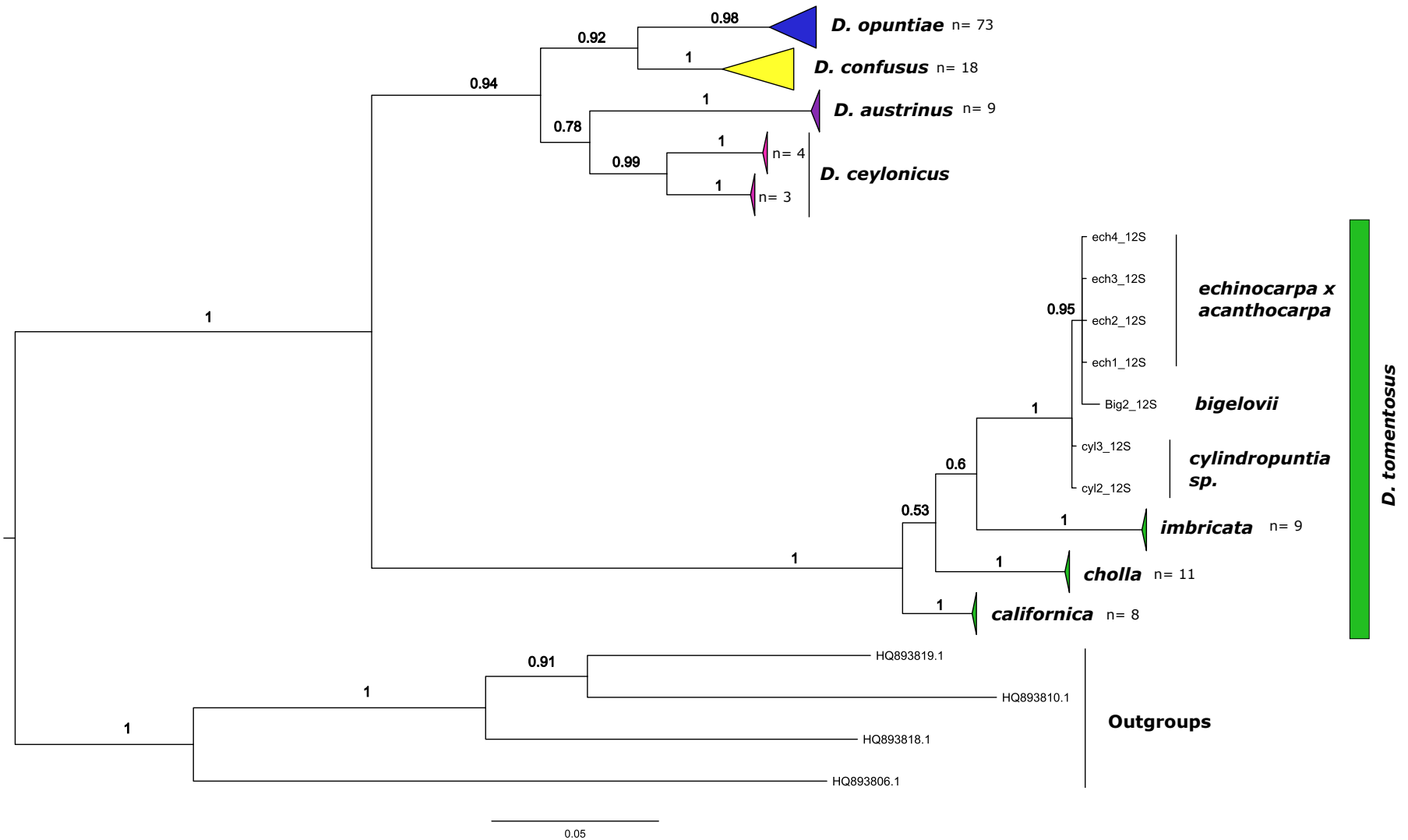
Extraction ID	Species	Lineage	Locality	Host plant
VS121O.4PM	<i>D. confusus</i>		USA: Arizona, Four Peaks Mt.	<i>O. engelmannii</i>
VS124O.lc	<i>D. confusus</i>		USA: New Mexico, Las Cruces	<i>O. engelmannii</i>
VS136	<i>D. confusus</i>		USA: Texas, Laredo	<i>O. engelmannii</i>
VS137	<i>D. confusus</i>		USA: Texas, Laredo	<i>O. engelmannii</i>
VS138	<i>D. confusus</i>		USA: Texas, Laredo	<i>O. engelmannii</i>
VS139	<i>D. confusus</i>		USA: Texas, Laredo	<i>O. engelmannii</i>
VS141	<i>D. confusus</i>		USA: Arizona, Tucson	<i>O. engelmannii</i>
VS142	<i>D. confusus</i>		USA: Arizona, Tucson	<i>O. engelmannii</i>
VS143	<i>D. confusus</i>		USA: New Mexico, Las Cruces	<i>O. engelmannii</i>
VS144	<i>D. confusus</i>		USA: New Mexico, Las Cruces	<i>O. engelmannii</i>
VS110	<i>D. opuntiae</i>	'ficus'	South Africa: EC, Prieel, near Fort Beaufort	<i>O. ficus-indica</i>
VS111	<i>D. opuntiae</i>	'ficus'	South Africa: EC, Prieel, near Fort Beaufort	<i>O. ficus-indica</i>
VS112	<i>D. opuntiae</i>	'ficus'	South Africa: EC, Prieel, near Fort Beaufort	<i>O. ficus-indica</i>
VS113	<i>D. opuntiae</i>	'ficus'	South Africa: EC, Prieel, near Fort Beaufort	<i>O. engelmannii</i>
VS114	<i>D. opuntiae</i>	'ficus'	South Africa: EC, Prieel, near Fort Beaufort	<i>O. engelmannii</i>
VS115	<i>D. opuntiae</i>	'ficus'	South Africa: EC, Prieel, near Fort Beaufort	<i>O. engelmannii</i>
VS116	<i>D. opuntiae</i>	'ficus'	South Africa: EC, Kluklu farm, near Fort Beaufort	<i>O. engelmannii</i>
VS117	<i>D. opuntiae</i>	'ficus'	South Africa: EC, Kluklu farm, near Fort Beaufort	<i>O. engelmannii</i>
VS118	<i>D. opuntiae</i>	'ficus'	South Africa: EC, Kluklu farm, near Fort Beaufort	<i>O. engelmannii</i>
VS119O.flag	<i>D. opuntiae</i>		USA: Arizona, Flagstaff	<i>O. engelmannii</i>
VS120O.flag	<i>D. opuntiae</i>		USA: Arizona, Flagstaff	<i>O. engelmannii</i>
VS126O.sed	<i>D. opuntiae</i>		USA: Arizona, Sedona	<i>O. engelmannii</i>
VS129	<i>D. opuntiae</i>		USA: Texas, Moor Airbase	<i>O. engelmannii</i>
VS130	<i>D. opuntiae</i>		USA: Texas, Port Isabel	<i>O. engelmannii</i>
VS131	<i>D. opuntiae</i>		USA: Texas, Port Isabel	<i>O. engelmannii</i>
VS132	<i>D. opuntiae</i>		USA: Texas, Laredo	<i>O. engelmannii</i>
VS133	<i>D. opuntiae</i>		USA: Texas, Laredo	<i>O. engelmannii</i>
VS134	<i>D. opuntiae</i>		USA: Texas, Webb	<i>O. engelmannii</i>
VS135	<i>D. opuntiae</i>		USA: Texas, Webb	<i>O. engelmannii</i>
VS140	<i>D. opuntiae</i>		USA: California, Rincon	<i>O. engelmannii</i>
VS145	<i>D. opuntiae</i>		USA: Texas, Pecos County	<i>O. engelmannii</i>
VS092Ti	<i>D. tomentosus</i>	'imbricata'	South Africa: Uitenhage MRF	<i>C. imbricata</i>
VS093Ti	<i>D. tomentosus</i>	'imbricata'	South Africa: Uitenhage MRF	<i>C. imbricata</i>
VS100Tf	<i>D. tomentosus</i>	'cholla'	South Africa: Uitenhage MRF	<i>C. fulgida</i> var. <i>mamillata</i>

**Table A3:** Genbank sequences used to supplement the dataset for this project. PcoF1 & LepR1 and DTOMf & HCO2198 are both COI primers that amplify the same region of the COI gene.

Genbank accession numbers				Organism
12S	18S	PcoF1 & LepR1	DTOMf & HCO2198	
HQ893819.1	KY9276001			<i>Parasaissetia nigra</i>
HQ893810.1	HQ8937881			<i>Ceroplastes sinensis</i>
HQ893818.1	HQ8937931	GU936953.1		<i>Coccus viridis</i>
KT199030.1	KT199030.1			<i>Marchalina hellenica</i>
HQ893806.1	KT199034.1	KP692469.1		<i>Icerya purchasi</i>
KJ701959.1	KJ7019951			<i>Dactylopius coccus</i>
KJ701960.1	KJ7019941			<i>Dactylopius coccus</i>
KJ701961.1	KJ7019931			<i>Dactylopius coccus</i>
KJ701962.1	KJ7019921			<i>Dactylopius coccus</i>
KJ701963.1	KJ7019911			<i>Dactylopius coccus</i>
			GU228789.1	<i>Ceroplastes sp.</i>
			GU228788.1	<i>Dactylopius opuntiae</i>
			GU228787.1	<i>Dactylopius tomentosus</i> 'Imbricata'
			GU228786.1	<i>Dactylopius tomentosus</i> 'Imbricata'
			GU228781.1	<i>Dactylopius tomentosus</i> 'RoseaP'
			GU228781.1	<i>Dactylopius tomentosus</i> 'ChollaDH'
			GU228780.1	<i>Dactylopius tomentosus</i> 'ChollaE'
			GU228779.1	<i>Dactylopius tomentosus</i> 'ChollaD'
			GU228785.1	<i>Dactylopius tomentosus</i> 'RoseaY'
			GU228784.1	<i>Dactylopius tomentosus</i> 'Fulgida'
			GU228783.1	<i>Dactylopius tomentosus</i> 'Mamallata'
			GU228782.1	<i>Dactylopius tomentosus</i> 'Molesta'
			AB440069.2	<i>Coccus hesperidum</i>
			AB440071.1	<i>Eulecanium kunoense</i>
			AB440073.1	<i>Parthenolecanium pruinosum</i>
			AB440072.2	<i>Pulvinaria psidii</i>

**Table A4:** Rhodes University and Agricultural Research Council (ARC) accession numbers for *Dactylopius* samples from the native range (USA), identified by Ian Millar in 2019. Representative sample IDs refer to those used in the gene phylogenies in this project. See Table A1 for the full list of sequences.

Rhodes University Accession Number	ARC Accession Number	Identity	Representative sample ID
RH1283	HC 7331	<i>D. confertus</i>	
RH1284	HC 7331	<i>D. confertus</i>	CSW001H
RH1285	HC 7331	<i>D. confertus</i>	
RH1286	HC 7332	<i>D. confusus</i>	VS072u
RH1287	HC 7333	<i>D. opuntiae</i>	VS083u
RH1288	HC 7334	<i>D. opuntiae</i>	VS069u
RH1289		<i>D. confusus</i>	VS080u
RH1290	HC 7335	<i>D. opuntiae</i>	VS064u
RH1291	HC 7336	<i>D. opuntiae</i>	VS077u
RH1292	HC 7337	<i>D. opuntiae</i>	VS083u
RH1293	HC 7338	<i>D. confusus</i>	VS073u
RH1294		<i>D. confusus</i>	VS078u
RH1295		<i>D. opuntiae</i>	VS066u



**Figure A3:** 12S rRNA Bayesian phylogeny based on the alignment obtained from R-Coffee, which aligns RNA sequences according to their secondary structure. The HKY + I + G model was used, where shapepr = 0.9980, pinvarp = 0.1420 and tratiopr = 1.4096. State frequencies were 0.5155, 0.0897, 0.0282 and 0.3665. Major clades have been collapsed.

## 5.3 Program settings and file formats

### 5.3.1 GeneMarker

1. Size standard set to 'GS500LIZ' (Size editor → File → Import ABI Size Standard → select the desired size standard file) and Analysis Type to 'AFLP'.
2. The selected Raw Data Analysis settings were: Auto Range (frame), Saturation Repair, Baseline Subtraction, Pull-up Correction, Spike Removal and Local Southern.
3. Allele Call settings were: Allele calls set to the 'automatic'. Minimum Intensity for peak detection set at 50 RFU (relative fluorescence units) and Max Intensity left at the default (30000 RFU). This minimum intensity is recommended by Arrigo *et al.* (2012) as a more permissive prior screening approach. Prince (2009) used an even lower threshold, set at 20 RFU. Peaks below the threshold are most likely due to the slippage of polymerase during the PCR elongation step, sometimes referred to as 'taq stutter' (AppliedBiosystems, 2014; Prince, 2009). Plus-A Filter was left selected.
4. 'AFLP normalization' selected.
5. The summary table was obtained by selecting 'Report settings' followed by 'Peak Table'. The 'Abide by Panel', 'Show Only Uncertain Alleles' and 'Show Rejected Low Score Alleles' options were deselected, and the \* symbol in the 'Show \* when no allele call is unchecked' option was deleted and left blank. Peak reads were subsequently converted into a summary table that was exported as a .txt file.

### 5.3.2 POPART

POPART reads in files in Nexus format, following the minimal working example below. In the 'begin traits' section, the trait label index number corresponds to the column number for each sequence (i.e. a '1' in the first column = Africa, in the second column = the USA and the third column = Asia), such that sequenceA and sequenceC represents Africa, sequenceB the USA and sequenceD Asia.

```
# NEXUS
begin data;
```

```

dimensions ntax=4 nchar=4;
format datatype=dna missing=N gap = -;
matrix
sequenceA   ACTG
sequenceB   ACGG
sequenceC   ACTT
sequenceD   ACCC
;
end;

begin traits;
Dimensions ntraits=3;
Format labels=yes separator=Comma;
TraitLabels Africa, USA, Asia;
matrix
sequenceA   1,0,0
sequenceB   0,1,0
sequenceC   1,0,0
sequenceD   0,0,1
;
end;

```

### 5.3.3 SplitsTree

```

#NEXUS
BEGIN Taxa;
DIMENSIONS NTAX = 2;
END;

BEGIN Characters;
DIMENSIONS NCHAR = 4;

```

```

FORMAT DATATYPE = STANDARD INTERLEAVE MISSING = ?;
MATRIX
Seq1 001?
Seq2 11??
;
END;

```

### 5.3.4 STRUCTURE

Binary data were organised in Microsoft Excel<sup>®</sup> according to the example format below. Question mark (“?”) symbols replaced with ‘-9’, saved as a tab-delimited .txt file, and uploaded to STRUCTURE. Column 2 contained integers specifying population groupings, referred to as ‘PopData’ in the STRUCTURE user manual. These groupings are not used as priors by default, and are otherwise useful to organise the graphical output produced by CLUMPAK (Kopelman *et al.*, 2015) (available at <http://clumpak.tau.ac.il/index.html>).

		Locus 1	Locus 2	Locus 3
Sample1	1	0	1	-9
Sample2	1	0	1	1
Sample3	2	1	1	1
Sample4	2	-9	0	0

Ploidy was set to 1, missing data to -9, ‘Row of marker names’ and ‘individual ID for each individual’ were selected, as well as ‘Putative population origin for each individual’.  $K$  was set from 1 through to at least four times the expected number of lineages to ensure thorough sampling. 20 iterations were run for each value of  $K$  to ensure reliable output, with a burn-in of 100 000, followed by 50 000 MCMC runs (following the recommendations made by Wang (2017) and Puechmaille (2016)). Puechmaille (2016) found that increasing the burn-in and/or MCMC runs above these values had no further effect on the result. The ‘Population admixture model’ and ‘Allele frequencies correlated’ were selected (recommended by Falush *et al.* (2003) and tested by Puechmaille (2016)). All other parameters were left at their default

settings (similar to the approach made by Sutton *et al.* (2017), and Medrano *et al.* (2014)).

## 5.4 Jaccard's transformation

Seq x	1	0	1	0	0	1	1	0	1	?
Seq y	1	1	1	0	0	1	0	1	1	1

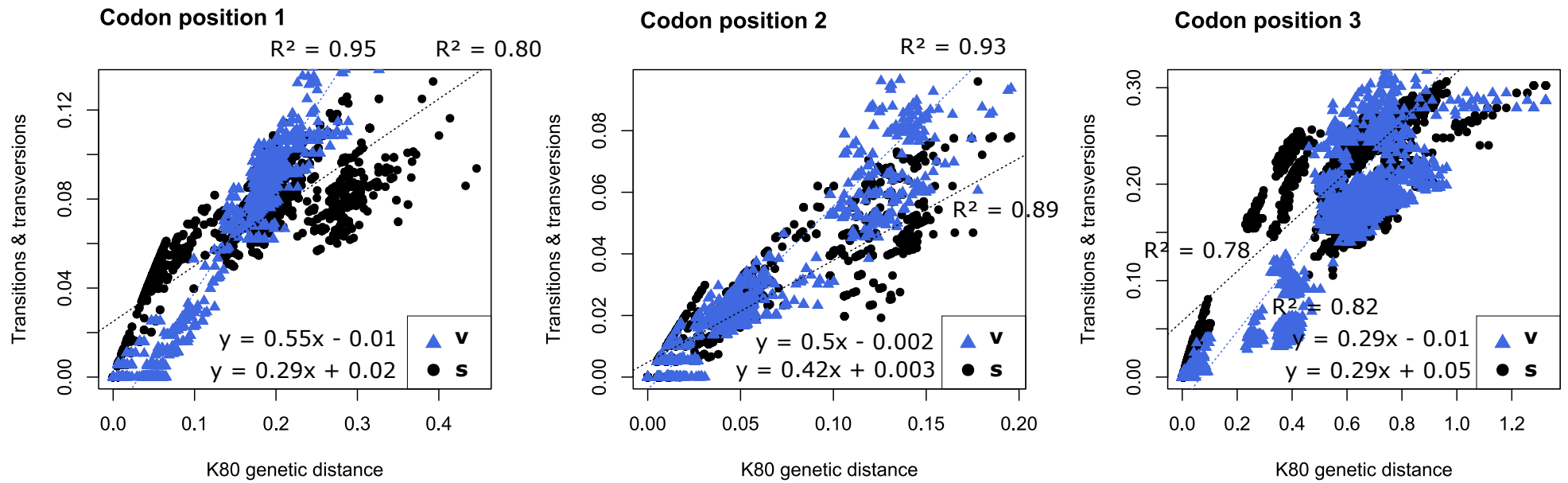
$$\begin{aligned}
 J_i &= f_{11}/(f_{01} + f_{10} + f_{11}) \\
 &= 4/(2 + 1 + 4) \\
 &= 0.57
 \end{aligned}$$

**Figure A4:** An example of Jaccard's similarity index calculation. Note that the shared absence of bands are not taken into account in the calculation of pairwise distances. Jaccard's distance = 1 -  $J_i$ , and therefore 0.43 in this case.

## 5.5 Codon saturation

**Table A5:** Codon saturation results from DAMBE for the protein-coding COI-A and COI-B gene regions. A significant p-value and an Iss value < the Iss.cSym value means that there is little saturation present.

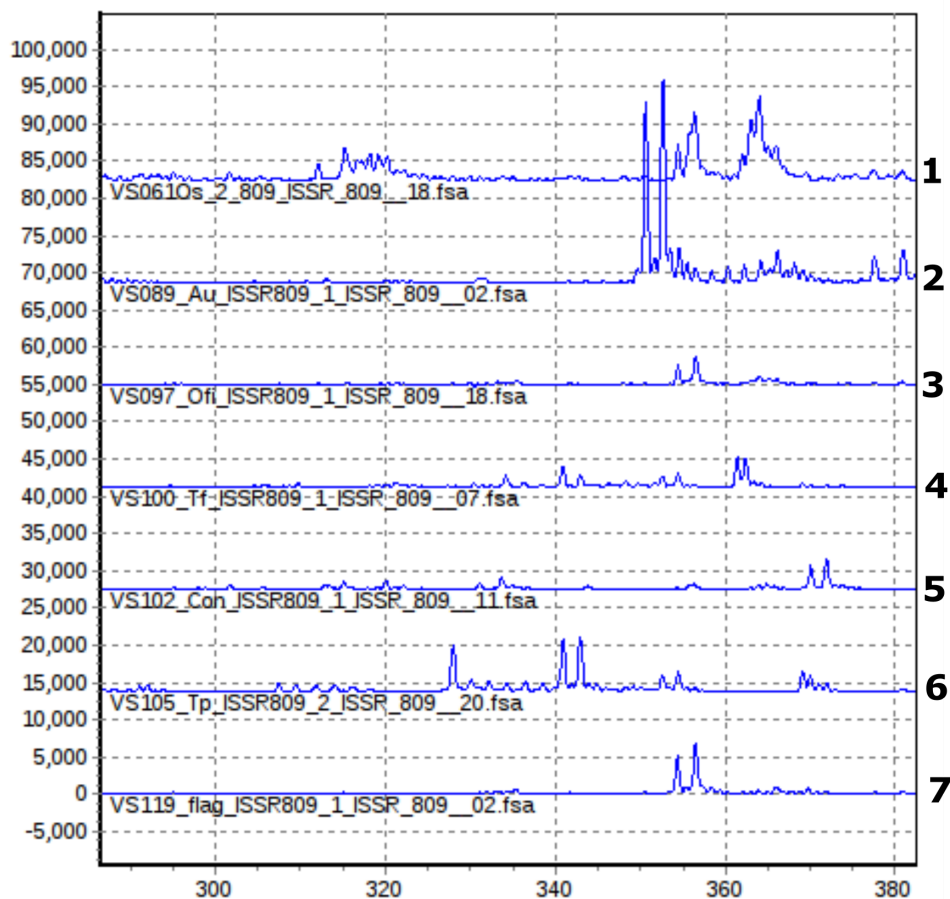
	NumOTU	Iss	Iss.cSym	T	DF	P
Codon Position 1	4	0.357	0.777	10.403	200	0
	8	0.327	0.734	10.155	200	0
	16	0.311	0.643	8.428	200	0
Codon Position 2	4	0.273	0.777	12.203	200	0
	8	0.227	0.734	12.727	200	0
	16	0.228	0.643	10.415	200	0
Codon Position 3	4	0.578	0.777	4.956	200	0
	8	0.539	0.734	4.979	200	0
	16	0.534	0.643	2.843	200	0
Codon Position 1 & 2	4	0.334	0.788	14.989	401	0
	8	0.274	0.742	16.002	401	0
	16	0.260	0.702	15.465	401	0



**Figure A5:** Saturation graphs for codon positions 1 to 3 for the COI gene region; showing the relationship between genetic distance and the rate of transitions (s) and transversions (v).

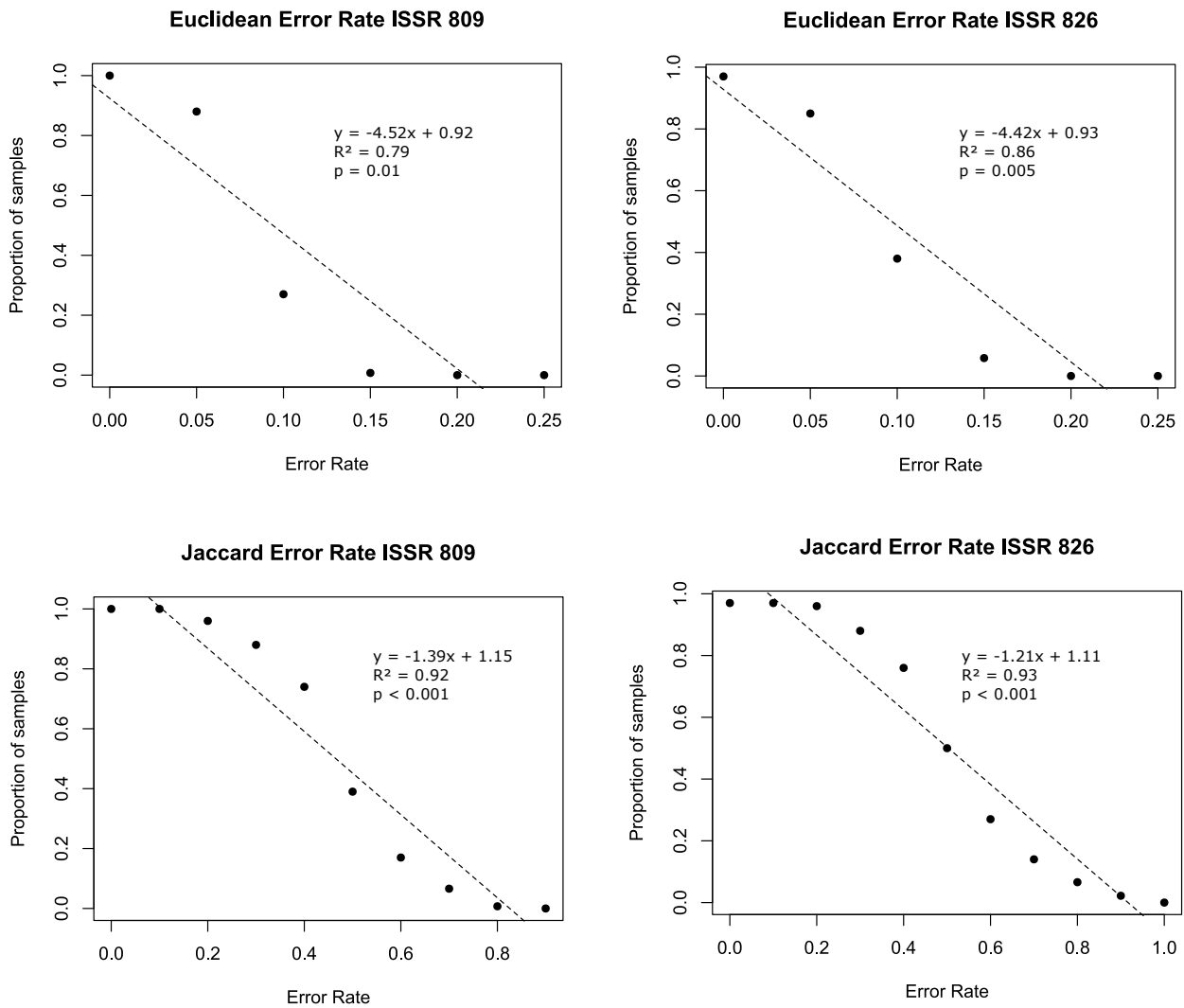
## 5.6 ISSR results

### 5.6.1 Electropherograms

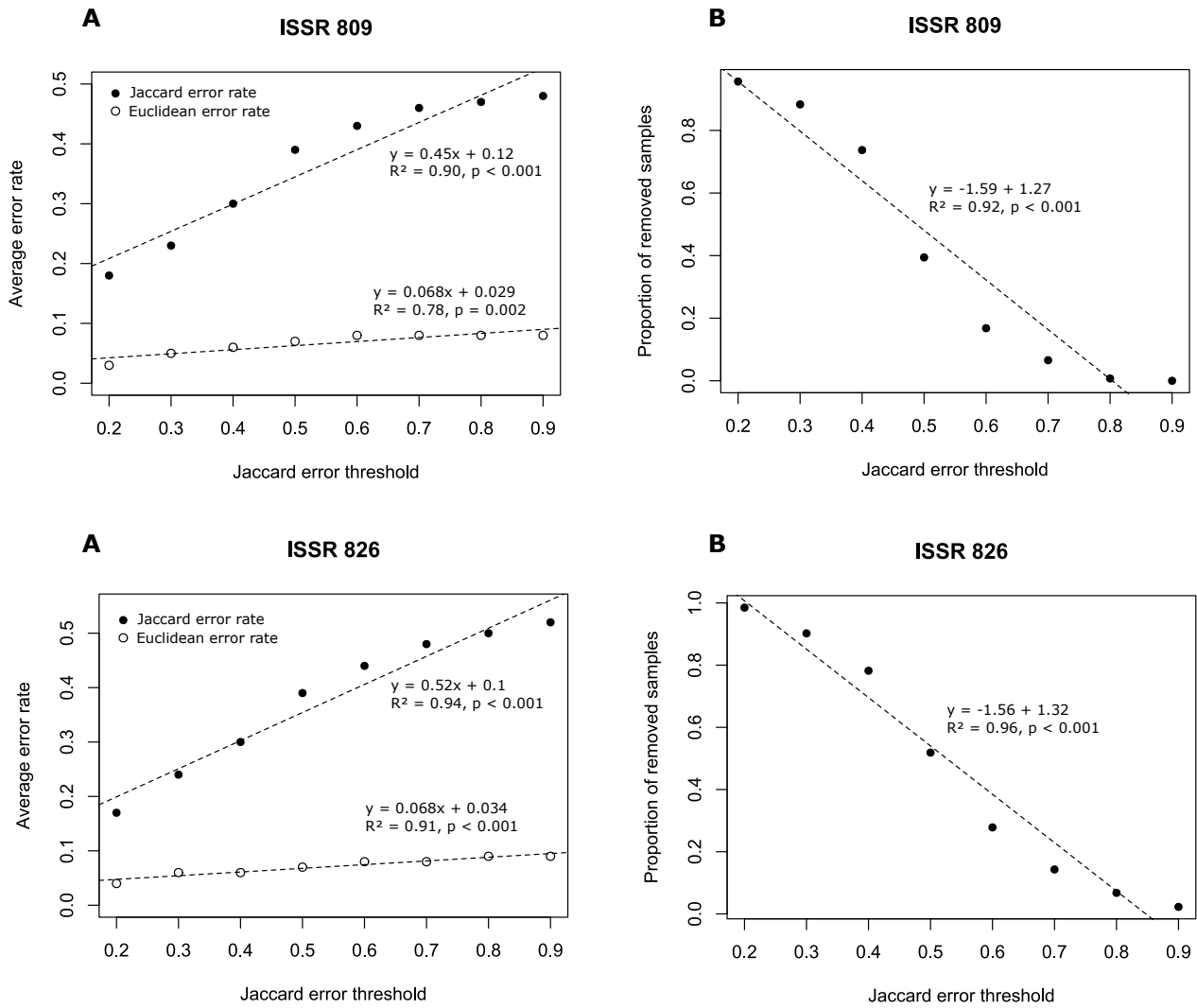


**Figure A6:** An overlay view of fragment sizes between 300 and 380 bp for samples (1) *Dactylopius opuntiae* 'stricta', (2) *D. austrinus*, (3) *D. opuntiae* 'ficus-indica', (4) *D. tomentosus* 'cholla', (5) *D. confusus*, (6) *D. tomentosus* 'pallida' and (7) *D. opuntiae* collected from Flagstaff (AZ, USA). The y-axis represents Relative Fluorescence Units (RFU), and the x-axis displays fragment size in base pairs.

## 5.6.2 Error rates

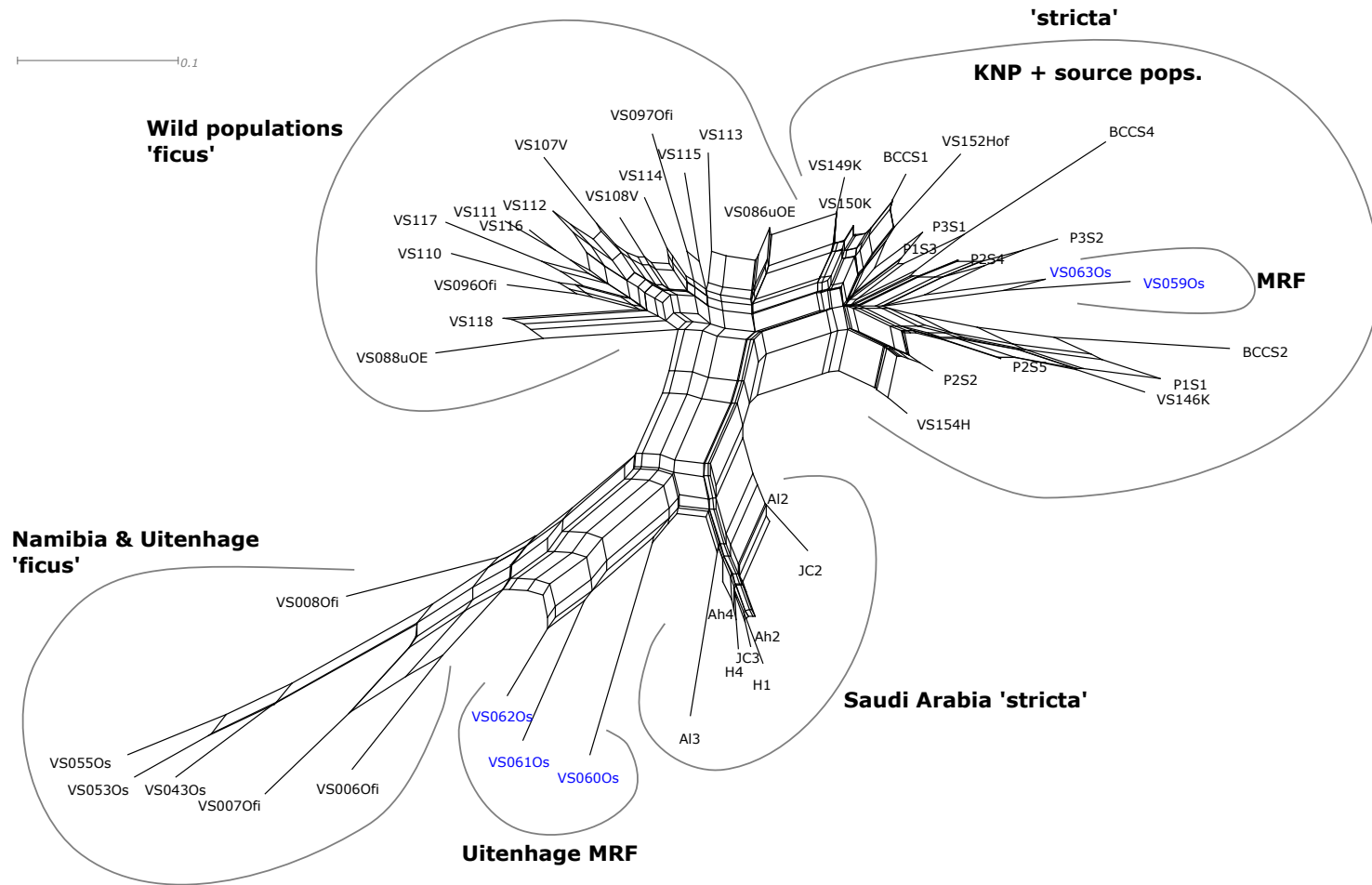


**Figure A7:** Euclidean and Jaccard error rates for ISSR primers 826 and 809, showing the proportion of samples greater than a particular error rate. Values are representative of the full data set.

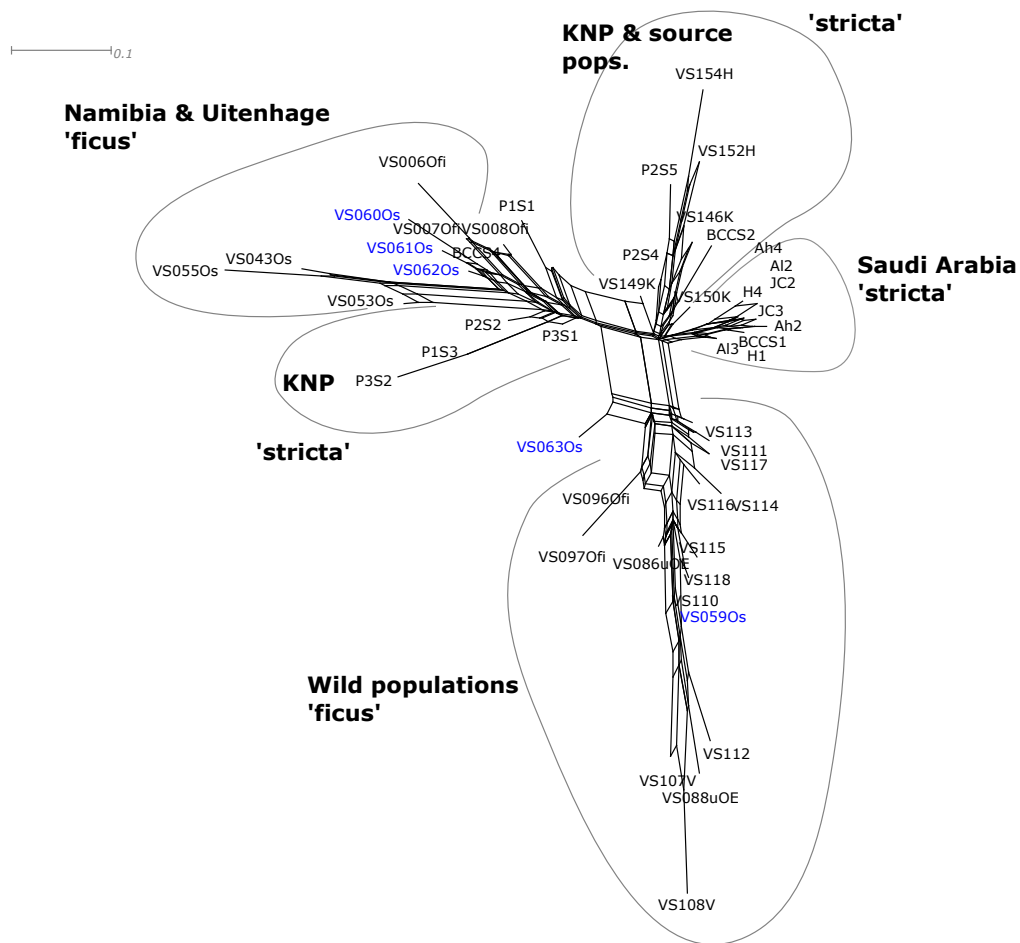


**Figure A8:** Graphical representations of the effect of the removal of samples (ISSR 809 and ISSR 826) with a Jaccard error rate greater than a specific threshold on A) the average Jaccard and Euclidean error rates and B) the proportion of removed samples from the data set.

### 5.6.3 Splitstree

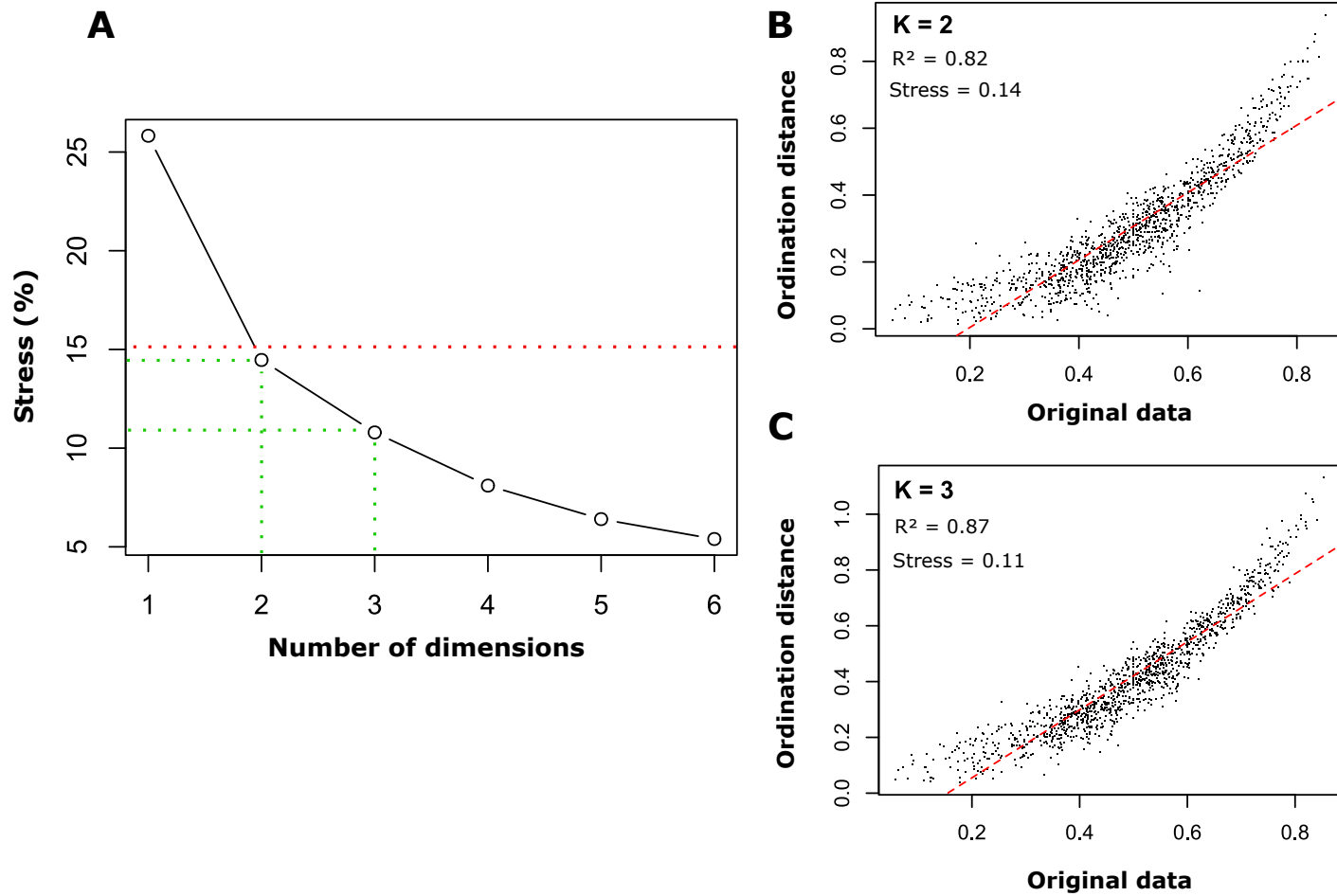


**Figure A9:** Splitstree diagram representative of ISSR 809 only. Mass Rearing Facility (MRF) samples are shown in blue. KNP = Kruger National Park, and source pops. = the 'stricta' culture kept by Hildegard Klein and John Hoffmann.



**Figure A10:** Splittree diagram representative of ISSR 826 only. Mass Rearing Facility (MRF) samples are shown in blue. KNP = Kruger National Park, and source pops. = the 'stricta' culture kept by Hildegard Klein and John Hoffmann.

### 5.6.4 Shepard and scree plots



**Figure A11:** A) Scree plot showing the relationship between the number of dimensions used for a non-metric MDS, and the corresponding stress score. Stress scores below 15% are considered to be a good fit for the data (indicated by the dashed red line). Stress scores for  $K = 2$  and  $K = 3$  are indicated by the dashed green lines. Shepard plots are shown in B) ( $K = 2$ ) and C) ( $K = 3$ ), plotting the relationship between the original distance data and the distances resulting from the ordination transformation. A high  $R^2$  value is desirable, as that indicates an accurate reflection of the data in the ordination plot.

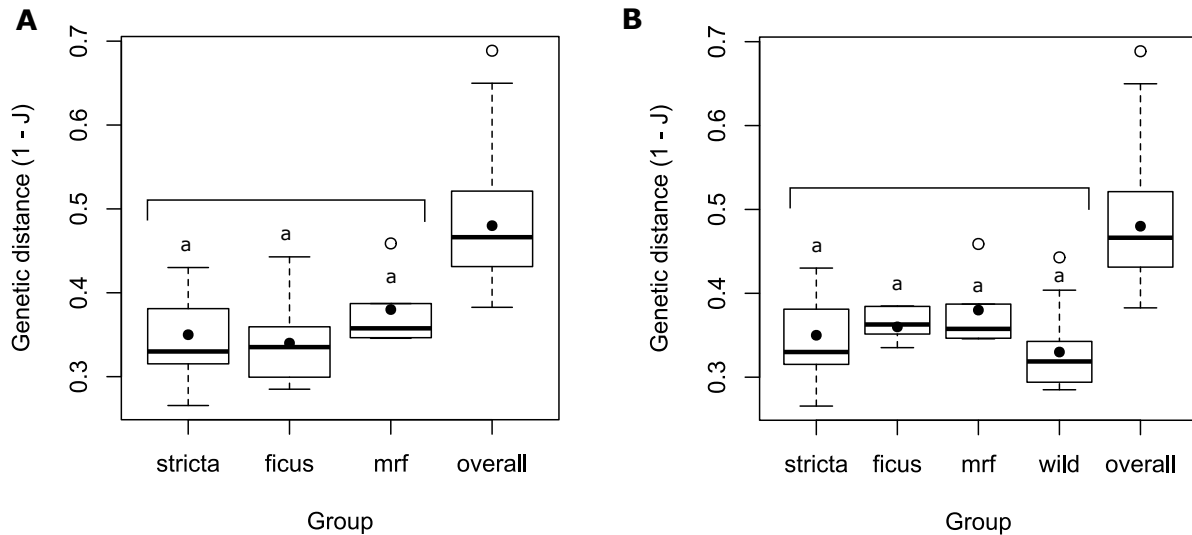
### 5.6.5 ISSR primer statistics

**Table A6:** POPGENE statistical output for each individual primer and their concatenation for the ‘stricta’ and ‘ficus’ genetic clusters. na = number of alleles, ne = effective number of alleles, h = Nei’s genetic diversity, I = Shannon’s information index, No. PM loci = number of polymorphic loci, % PM loci = percentage polymorphic loci, Ht = total genetic diversity, Hs = intra population genetic diversity, Gst = coefficient of gene differentiation. All values presented are averages. For the intra-group statistics, Saudi Arabia + Kruger N. Park + Klein & Hoffmann = ‘stricta’, and Uitenhage + Namibia + wild populations = ‘ficus’.

Primer	Group	na	ne	h	I	No. PM loci	% PM loci	Ht	Hs	Gst	
ISSR 809	stricta	1.20	1.09	0.06	0.09	78	20.31	0.06	0.03	0.51	
	ficus	1.41	1.17	0.11	0.17	156	40.62	0.14	0.06	0.59	
	Mass Rearing Facility	1.19	1.14	0.08	0.11	74	19.27				
	All	1.45	1.15	0.10	0.16	174	45.31	0.13	0.10	0.23	
	<b>INTRA-GROUP:</b>										
	Saudi Arabia	1.08	1.05	0.03	0.04	30	7.81				
	Kruger N. Park	1.09	1.06	0.03	0.05	35	9.11				
	Klein & Hoffmann	1.07	1.05	0.03	0.04	27	7.03				
	Uitenhage	1.15	1.13	0.07	0.10	56	14.58				
	Namibia	1.12	1.11	0.06	0.08	47	12.24				
Wild populations	1.15	1.08	0.05	0.07	58	15.1					
ISSR 826	stricta	1.03	1.03	0.01	0.02	11	3.01				
	ficus	1.03	1.03	0.02	0.02	12	3.28				
	stricta + ficus	1.06	1.04	0.03	0.04	23	6.28				

### 5.6.6 Genetic distance and diversity

All *D. opuntiae* specimens showed an average Jaccard distance of 0.48 (SE  $\pm$  0.01) (Fig. A12). None of the within-group genetic distances showed significant differences (Kruskal-Wallis ANOVA: Fig. A12A) Chi-square = 2.84, p= 0.24, and Fig. A12B) Chi-square = 6.72, p= 0.081). By group, ‘ficus’ had the highest percentage of polymorphic loci (PM) (38.53 %), as well as genetic diversity (h = 0.10, I = 0.16), while samples from the MRF had the lowest percentage PM loci (17.33 %) (Table A7).



**Figure A12:** Box plots for within and overall group genetic distances for *D. opuntiae*, measured by the Jaccard distance (1-J). A) ‘ficus’ taken as a whole and B) wild populations removed from ‘ficus’ and put into its own group. Solid black dots are the means, and the same letters above each box indicate non-significance.

**Table A7:** POPGENE statistical output for the concatenation of ISSR 809 and ISSR 826 for the ‘stricta’ and ‘ficus’ genetic clusters. na = number of alleles, ne = effective number of alleles, h = Nei’s genetic diversity, I = Shannon’s information index, No. PM loci = number of polymorphic loci, % PM loci = percentage polymorphic loci, Ht = total genetic diversity, Hs = intra population genetic diversity, Gst = coefficient of gene differentiation. All values presented are averages. For the intra-group statistics, Saudi Arabia + Kruger N. Park + Klein & Hoffmann = ‘stricta’, and Uitenhage + Namibia + wild populations = ‘ficus’.

Primer	Group	na	ne	h	I	No. PM loci	% PM loci	Ht	Hs	Gst	
ISSR 809 & ISSR 826	stricta	1.22	1.10	0.06	0.10	165	22	0.06	0.03	0.48	
	ficus	1.39	1.16	0.10	0.16	289	38.53	0.14	0.05	0.63	
	Mass Rearing Facility	1.17	1.13	0.07	0.10	130	17.33				
	All	1.45	1.15	0.10	0.16	337	44.93	0.12	0.10	0.21	
	<b>INTRA-GROUP:</b>										
	Saudi Arabia	1.07	1.05	0.03	0.04	55	7.33				
	Kruger N. Park	1.13	1.07	0.04	0.07	98	13.07				
	Klein & Hoffmann	1.07	1.05	0.03	0.04	50	6.67				
	Uitenhage	1.10	1.09	0.05	0.07	78	10.40				
	Namibia	1.12	1.10	0.05	0.08	86	11.47				
	Wild populations	1.15	1.08	0.05	0.07	111	14.8				

## 5.7 R Scripts

### 5.7.1 Barcode testing using SPIDER

```

1 # Read in a .fas file
2 setwd("file_location")
3 getwd()
4 input = "file_name.fas"

```

```

5 seqs = read.dna(input, format = 'fas')
6 dist = dist.dna(seqs, pairwise.deletion = TRUE)
7 # Nearest neighbour:
8 sppNames = dimnames(seqs)
9 table(nearNeighbour(dist, sppNames))
10 # Threshold Identification:
11 table(threshID(dist, sppNames, threshold = 0.01))
12 # Meier's Best Close Match:
13 table(bestCloseMatch(dist, sppNames, threshold = 0.01))
14 # Optimising the threshold value from the default of 0.01:
15 threshVal = seq(0.001,0.02, by = 0.001)
16 sens = lapply(threshVal, function(x) threshOpt(dist, sppNames, thresh = x))
17 sensMat = do.call(rbind, sens)
18 barplot(t(sensMat)[4:5,], names.arg=paste((sensMat[,1]*100), "%"))
19 # Graphically displaying the barcoding gap:
20 inter = nonConDist(dist, sppNames) *100
21 intra = maxInDist(dist, sppNames) *100
22 bnd = cbind(data.frame(inter, intra))
23 ord = bnd[order(bnd$inter),]
24 plot(ord$inter, type="n", ylab="Percent K2P distance", xlab="Individual")
25 segCol = rep("gray50", length(ord$inter))
26 segCol[ord$inter-ord$intra < 0] = "red"
27 segments(x0=1:length(ord$inter), y0=ord$inter, y1=ord$intra, col=segCol, lwd=6)

```

## 5.7.2 Jaccard's transformation

```

1 binarymat <-read.csv("file_name.csv", header = T, sep = ",", dec = ".")
2 # removes the first two columns, which contain sample names and grouping information
3 binarymat2 = binarymat[,c(-1,-2)]

```

```

4 # assign sample names as row names
5 row.names(binarymat2) = binarymat[,1]
6 # shortens the names to 10 characters
7 newnames =substring(row.names(binarymat2), 0, 10)
8 row.names(binarymat2) = newnames
9 # replaces ? with NA, as NA's are ignored.
10 binarymat2[binarymat2=="?"] <- NA
11 # converts the matrix to a data frame
12 binarymat2 = as.data.frame(binarymat2)
13 # creates a distance matrix
14 d = dist((binarymat2), method = "binary", diag = TRUE, upper = T)
15 d = as.data.frame(as.matrix(d))
16 d2 = as.dist(d)

```

### 5.7.3 UPGMA clustering tree

```

1 library(pvclust)
2 # 'average' refers to the UPGMA method.
3 # binarymat2 is transformed such that the data is organised by column.
4 result <- pvclust(t(binarymat2),
5   ↪ method.dist = "binary", method.hclust="average", nboot=1000)
6 p = plot(result, cex = 0.6, print.num = F, print.pv = T, cex.pv = 0.6)

```

### 5.7.4 Non-metric MDS plot

```

1 # k is the number of dimensions
2 # d2 is the distance matrix obtained for the data using the `binary' method
3 fit <- isoMDS(d2, k=2)
4 # view the stress value

```

```

5 fit$stress
6 mds_plot = isoMDS(d2,k=2)
7 eqscplot(mds_plot$points, col = c("red", "dodgerblue3", "forestgreen",
  ↪ "orange")[binarymat$Group], pch = c(15, 16, 17,18)[binmat$Group])
8 # the number of colours and pch
  ↪ types selected will depend on the number of groups (binarymat$Group) inputted
9 scree.plot = function(d, k) {
10 stresses=isoMDS(d, k=k)$stress
11 for(i in rev(seq(k-1)))
12 stresses=append(stresses,isoMDS(d, k=i)$stress)
13 plot(seq(k),rev(stresses),
  ↪ type="b", xaxp=c(1,k, k-1), ylab="Stress", xlab="Number of dimensions")
14 }
15 # plot stress values for six dimensions
16 scree.plot(d2, k=6)
17 shep<-Shepard(d2, mds_plot$points, p=2)
18 plot(shep, pch=".", main = "K = 2")
19 # R squared value
20 summary(lm(shep$y~shep$x))
21 # line of best fit through the points
22 abline(lm(shep$y~shep$x), col = "red", lty = 2)

```

### 5.7.5 Three dimensional non-metric MDS plot

Although not included in the results, three dimensional nMDS (Non-metric Multidimensional Scaling) scatter plots were created in R using the plot3D package (Soetaert, 2017). The following following code was implemented:

```

1 # nMDS with three dimensions

```

```

2 fit <- isoMDS(d2, k=3)
3 x <- fit$points[,1]
4 y <- fit$points[,2]
5 z <- fit$points[,3]
6 library(plot3D)
7 # points are coloured by group, which was column 2 in the original binarymat matrix
8 scatter3D(x,y,z, phi = 25, type = "h",
9 pch = 19, cex = 1.5, bty = "b", colvar = as.integer(binarymat$Group),
10 ↪ col = c("blue", "red"), colkey = list(side = 2, length = 0.5))
11 # an optional addition to add sample names above points in the scatter plot
12 text3D(x,
13 ↪ y, z, labels = rownames(binarymat2), add = TRUE, colkey = FALSE, cex = 0.7)

```

## 5.7.6 ANOSIM and PERMANOVA

```

1 # ANOSIM:
2 anosim(d2, binarymat$Group, permutations = 999, distance = "jaccard")
3 # PERMANOVA:
4 library(vegan)
5 adonis(d2~binarymat$Group)
6 library(RVAideMemoire)
7 pairwise.perm.manova(d2, binarymat$Group)

```

## 5.8 Shiny Applications

**GENETIC BARCODE TESTER**

Created by Clarke van Steenderen

Home: Barcode testing | Best Close Match | Nearest Neighbour | Threshold ID | Barcode Gap | Rosenberg's Test of Monophyly | Find Optimal Threshold Value

Display Distance Matrix | PLOT BASE FREQUENCIES | PLOT GC CONTENT | CLUSTERING TREE (fragment data)

**1** The user can upload nucleotide sequences or binary data

Select data type: Nucleotide sequences, ISSR/FLP fragment analysis data

**2** User manual

**GENETIC BARCODE TESTER USER GUIDE**

Created by: Clarke van Steenderen  
 Based upon the algorithms derived from the R package SPIDER  
 Department of Zoology and Entomology  
 The Centre for Biological Control  
 Rhodes University, Grahamstown, Eastern Cape, South Africa  
 2019  
 e-mail: vsteenderen@gmail.com

**OVERVIEW**

This program enables the user to test how efficient a set of genetic sequences are in allowing for their correct identification. The barcode testing algorithms implemented here are derived from the R package SPIDER, (SPecies IDentity and Evolution in R) developed by Brown *et al.* (2012). The program has been encoded to allow for both nucleotide and binary data input (such as ISSRs and AFLPs), which are subsequently converted into distance matrices.

A highly useful function is that of the setting of distance thresholds in order to minimise the occurrence of false positive and false negative identifications. This attempts to help answer the question of "How different must the specimens be before they are considered different species/lineages/haplotypes etc.?".

Evolutionary model: K80

Gamma shape (alpha): 0

Pairwise deletion

Type in your outgroups:

**3** LOAD DATA

**4** Selection of tests: Best Close Match, Nearest Neighbour, Threshold ID, Barcode Gap, Rosenberg's Test, Find the Optimal Threshold Value, Display Distance Matrix

**2** After the data is uploaded, the sequence names appear in a table. These need to be edited such that all the same species/taxa of interest have the same name.

	row.names(data1)
1	ech
2	ech
3	ech
4	ech
5	cyl
6	cyl
7	chol
8	chol
9	chol
10	chol
11	cal
12	cal
13	cal
14	cal
15	big
16	big

**Figure A13:** Genetic barcode tester GUI showing the components of the application. Numbers in circles indicate the order in which data is uploaded and processed.

1  
2  
3

4  
5  
6

7  
8  
9

10  
11  
12

13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100

101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200

# BINMAT: FOR FRAGMENT ANALYSIS DATA

Created by Clarke van Steenderen

Click the ? to view the help file: ?

What type of data do you have?

The whole dataset are replicates for one species/location

Each pair in the dataset are replicates

Select a .csv file

Browse... No file selected

Preview

Check my data for unwanted values

Consolidate matrix

Download Matrix

MATRIX SUMMARY ERROR RATES UPGMA TREE nMDS PLOT nMDS Validation

	L1	L2	L3	L4	L1.1	L2.1	L3.1	L4.1	L1.2	L2.2	L3.2	L4.2
Sample 1+Sample 2	?	?	1	0	?	?	1	0	?	?	1	0
Sample 3+Sample 4	1	1	?	0	1	1	?	0	1	1	?	0
Sample 5+Sample 6	?	?	1	0	?	?	1	0	?	?	1	0
Sample 7+Sample 8	1	1	?	0	1	1	?	0	1	1	?	0
Sample 9+Sample 10	?	?	1	0	?	?	1	0	?	?	1	0
Sample 11+Sample 12	1	1	?	0	1	1	?	0	1	1	?	0

MATRIX SUMMARY ERROR RATES UPGMA TREE nMDS PLOT nMDS Validation

Summary info

Summary

Average no. peaks = 4.5

sd = 1.6432

Max. no. peaks = 6

Min. no. peaks = 3

No. loci = 12

Download Summary

Check Error rates

Errors

Average Euclidean Error: 0.375

Euclidean error St. dev: 0.1369

Average Jaccard: 0.5

Jaccard error St.dev: 0.1826

Download Errors

Remove samples with a jaccard error greater than:

0

Remove

Non-metric MDS plot tab:

	Groups	Colour_by_Group	Point_Shape
1	1	dodgerblue	16
2	2	black	16
3	3	red	16
4	4	green3	16
5	5	orange	16
6	6	darkblue	16
7	7	gold2	16
8	8	darkgreen	16
9	9	darkred	16

Remove samples with peaks less than:

0

Select a .csv file

Browse... GRAPES for CLARKE .csv

Upload complete

No. of dimensions (k):

2

Plot nMDS Download MDS Plot

Download filtered data set Download removed samples

Point size

0.1 2 5

Show sample labels?

No

Yes

Figure A14: Binary matrix processing application ('BINMAT'), where the user uploads a .csv file containing binary data replicates in order to consolidate them. Summary statistics, error rates and the construction of a UPGMA clustering tree are also available. Non-metric MDS plots can be created by uploading a consolidated binary matrix, where filtering of samples with a low peak count can be applied.

### 5.8.1 Dacty-ID

When uploading this application to the Shiny server, the following command was required in R in order to set the repositories to access packages within Bioconductor (namely the DECIPHER package by Wright (2016)):

```
setRepositories()
```

This produced the output on the console:

```
1: + CRAN
2: BioC software
3: BioC annotation
4: BioC experiment
5: CRAN (extras)
6: Omegahat
7: R-Forge
8: rforge.net
```

```
Enter one or more numbers separated by spaces, or an empty line to cancel
```

```
1:
```

Options 2, 3, and 4 were entered to include all the BioC-related dependencies.

# DACTYLOPIUS IDENTIFIER

Created by Clarke van Steenderen

**1** The user can choose from 4 gene databases, or upload their own in FASTA format

Select File

- Own file upload
- 12S.fas
- 18S.fas
- COI(DTOM).fas
- COI(PCOFI).fas
- Own file upload

**User manual** ?

## DACTYLOPIUS IDENTIFIER

### USER GUIDE

Created by:  
Clarke van Steenderen  
Department of Zoology and Entomology  
The Centre for Biological Control  
Rhodes University, Grahamstown, Eastern Cape, South Africa  
2019  
e-mail: vsteenderen@gmail.com

**OVERVIEW**

This program is designed for the quick identification of *Dactylopius* species (Hemiptera: Dactylopiidae), but can be applied to any taxon if the relevant database of genetic sequences is supplied. The user can either upload a database of their choice, or can select from those provided under the 'Select File' drop down menu (12S, 18S, and two COI markers). The sequences provided were obtained by Clarke van Steenderen as part of his M.Sc. degree, under the supervision of Dr. Iain Paterson and Dr. Shelley Edwards. A query nucleotide sequence is entered into the search box, and a phylogenetic tree is created where the query sequence is matched to the database, and highlighted in red.

**UPLOADING YOUR OWN FILE**

Click to view the help file: ?

Upload a .fas file

Browse... No file selected

Select File

Own file upload

Enter sequence name:

QUERY

Enter nucleotide sequence:

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

Distance correction method

Jukes-Cantor

Distance correction method

Jukes-Cantor

Agglomeration method

UPGMA

**3** Submit

Agglomeration method

- UPGMA
- complete
- single
- UPGMA
- WPGMA
- NJ
- ML

**2** Query sequence input

Distance correction method

Jukes-Cantor

Jukes-Cantor

**Query sequence shown in red. Its placement is relative to the selected database**

**Figure A15:** Identification GUI application, where the user can test the identity of a query sequence against a preselected genetic database. A phylogenetic tree is displayed on the screen, with the position of the query sequence highlighted in red.

## 5.9 MRBAYES Nexus files

### 5.9.1 12S

```
#NEXUS
[12S aligned sequences]

begin data;
dimensions ntax=152 nchar=779;
format interleave datatype=DNA missing=N gap=-;

matrix
ALIGNED 12S SEQUENCES
;
end;

begin mrbayes;
set autoclose=yes nowarn=yes;
lset nst=6 rates=gamma ngammacat = 4 code=universal;
outgroup HQ893819.1;
prset revmatpr = dirichlet(6.0485, 8.2345, 15.0465, 0.0004, 42.1397, 1.0000)
statefreqpr = dirichlet(0.3923, 0.1404, 0.1629, 0.3044)
shapepr = fixed(0.8960);
mcmc Ngen=50000000 printfreq=1000 samplefreq=1000 Nchains=4 savebrlens=yes starttree=random;
sumt burnin = 1250;
sump burnin = 1250;
end;
```

## 5.9.2 18S

```
#NEXUS
[18S aligned sequences]

begin data;
dimensions ntax=153 nchar=630;
format interleave datatype=DNA missing=N gap=-;

matrix
ALIGNED 18S SEQUENCES
;
end;

begin mrbayes;
set autoclose=yes nowarn=yes;
lset nst=6 rates=propinv ngammacat = 4 code=universal;
outgroup KY9276001;
prset   revmatpr   =   dirichlet(0.5109,      1.8724,      0.5109,      1.0000,      3.5607,
1.0000)
pinvarpr = fixed(0.6060);
mcmc   Ngen=50000000   printfreq=1000   samplefreq=1000   Nchains=4   savebrlens=yes   start-
tree=random;
sumt burnin = 1250;
sump burnin = 1250;

end;
```

### 5.9.3 COI

```
#NEXUS
[COI aligned sequences]

begin data;
dimensions ntax=97 nchar=603;
format interleave datatype=DNA missing=N gap=-;

matrix
ALIGNED COI SEQUENCES
;
end;

begin mrbayes;

set autoclose=yes nowarn=yes;
lset nst=2 rates= gamma ngammacat = 4 code=universal;
outgroup AB440072.2;
prset statefreqpr = dirichlet(0.3872, 0.1822, 0.0454, 0.3853)
shapepr = fixed(0.6260) tratioopr = fixed(2.1838);
mcmc Ngen=50000000 printfreq=1000 samplefreq=1000 Nchains=4 savebrlens=yes start-
tree=random;
sumt burnin = 1250;
sump burnin = 1250;

end;
```

# 6

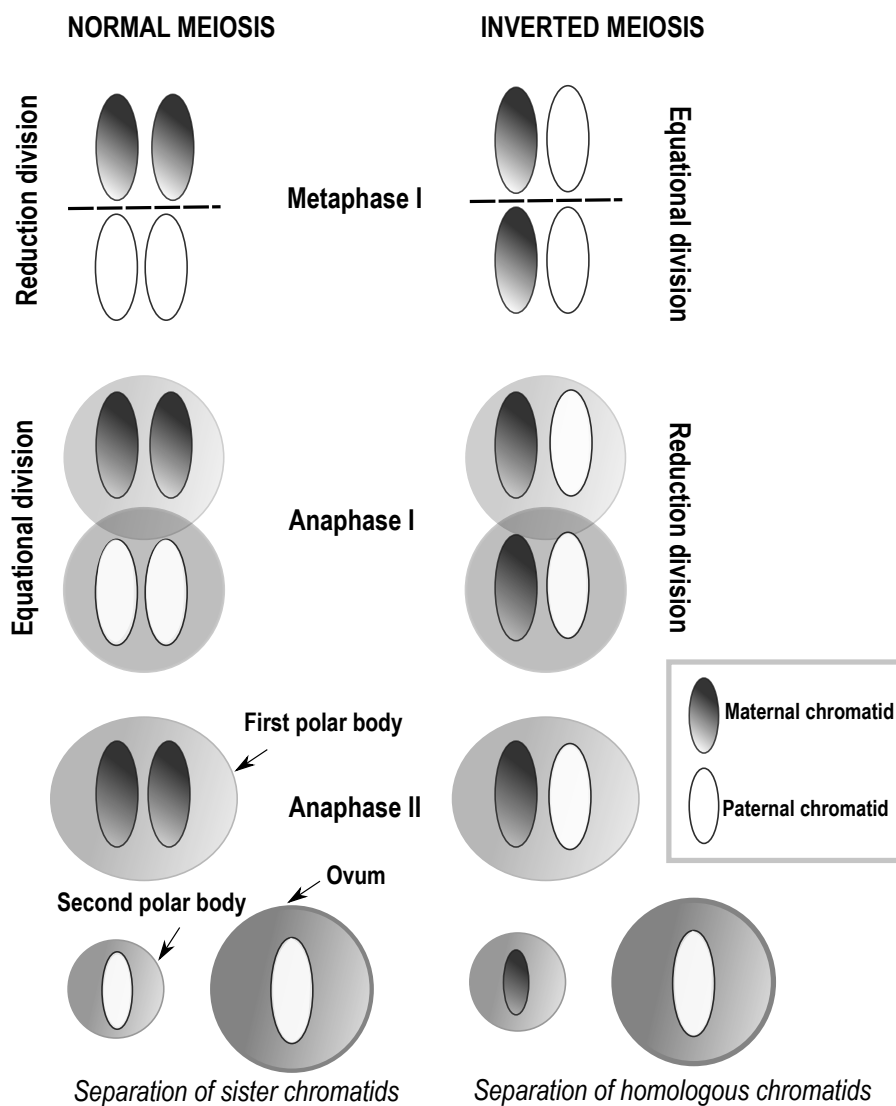
## Appendix B: Additional literature review material

In their hybrid study of *Dactylopius tomentosus* lineages, Mathenge *et al.* (2010b) postulated that male and female crawlers might display differing host specificities due to the Comstockiella genetic system displayed in this insect family. In other words, males are haploid, containing only a maternal genome, while females are diploid with both a maternal and paternal genome. Taking such underlying genetic processes into account may be useful in explaining observed trends in inherited traits and the resulting host-specificity displayed by hybrid offspring.

### 6.1 The genetic system of the scale insects

The scale insects display one of the most diverse genetic and sex-determination systems of all the insects, including diploidy, parthenogenesis, hermaphroditism and haplodiploidy (Burt and Trivers, 2009; Ross *et al.*, 2012). The ancestral genetic system of scale insects is that of XX-XO (where  $2n = 10$ ), such that females are homogametic, and males are heterogametic with only one sex chromosome (Hughes-Schrader, 1948; Nur, 1980). Arrhenotoky and paternal genome elimination (PGE) (also termed pseudoarrhenotoky or parahaplodiploidy) are variants of the haplodiploidy genetic system (Haig, 1993a; Gullan and Kosztarab, 1997; Normark, 2004a; Burt and Trivers, 2009; Ross *et al.*, 2010b). To further complicate matters, the Hemiptera possess holocentric chromosomes, where spindle fibres attach along the length of the chromosome during cell division (Hughes-Schrader, 1948). This is in contrast to the typical monocentric chromosomes where spindle fibres attach to a single centromere (Melters *et al.*, 2012). Holocentric chromosomes have allowed for a diverse range of karyotypes to form within the scale insects, such as species in the *Apiomorpha* genus (Coccoidea: Eriococcidae) that display chromosome numbers from  $2N = 4$  to  $2N = 92$  (Cook, 2000). This is because if holo-

centric chromosomes become fragmented, all segments are pulled to opposite poles of the cell by spindle fibres during meiosis, as opposed to monocentric chromosomes that do not include fragmented sections and thereby maintain their original karyotype (Melters *et al.*, 2012). Additionally, scale insects display inverse meiosis (Hughes-Schrader, 1948; Chandra, 1962; Bongiorni *et al.*, 2004). In this process, equational division occurs first followed by reduction division (Wrensch *et al.*, 1994; Bongiorni *et al.*, 2004). The difference between normal and inverse meiosis comes about due to the position of the paternal and maternal chromosomes relative to the orientation of the spindle fibres that form prior to metaphase I (Wrensch *et al.*, 1994) (Figure B1).



**Figure B1:** Difference between normal and inverted meiosis. Adapted from Brown and Nur (1964); Wrensch *et al.* (1994); Bongiorni *et al.* (2004)

### 6.1.1 The evolution of haplodiploidy

Haplodiploidy refers to a genetic system where males are haploid and contain only a maternal genome, and females are diploid with both a maternal and paternal genome (Normark, 2004a; de la Filia *et al.*, 2015). Haploidy in males occurs through either arrhenotoky (where females produce male offspring parthenogenetically) or PGE (where the paternal genome is eliminated at different developmental stages) (Bull, 1979; Burt and Trivers, 2009; Ross *et al.*, 2010b). Haplodiploidy is present in as many as 15% of arthropods (de la Filia *et al.*, 2015), and has evolved at least 10 times in insects (Normark, 2004a, 2006). Why this system evolved is not well understood, but the leading hypotheses suggest the antagonistic co-evolution (intragenomic conflict) between the sexes, conflict over sex-determination, the clearance of deleterious mutations and rapid adaptation, and the role of endosymbionts (intergenomic conflict) in determining chromosome dynamics are possible drivers (Normark, 2004a; Burt and Trivers, 2009; Ross *et al.*, 2010b; de la Filia *et al.*, 2015). Bull (1979) and Herrick and Seger (1999) suggested that maternal and paternal chromosomes engage in an arms race during spermatogenesis to transmit their genes to the next generation, and that this may explain the evolution of the different variations of PGE in scale insects. Females that produce haploid sons increase the transferal of their genes by two-fold in the gametes that their sons will produce (Bull, 1979). Normark (2006) suggested that if an insect group displays gregariousness (as scale insects do), maternal care or other close kin associations, and relatedness asymmetry (i.e. offspring are linked more closely through their maternal line than through paternal genes), then offspring are more likely to display cooperative rather than competitive behaviour towards each other. In this case, some male genes may pose as selfish elements and can decrease overall fitness.

The sex determination system of scale insects that display PGE is poorly understood, with no record of the observation of sex chromosomes in species with this genetic system (Brown and Nur, 1964; Ross *et al.*, 2010b). It is hypothesised that sex is determined by mechanisms such as gamete imprinting, the addition of maternal proteins to eggs (Ross *et al.*, 2010b), the activity of endosymbionts (Buchner, 1965), and factors such as temperature (Nelson-Rees, 1962), population density, and the age of females at mating and oviposition (Ross *et al.*, 2010a). Studies by Sabour (1972) and Bongiorno *et al.* (2007) suggest that PGE and sex determination are maternally controlled through the production of histone proteins such as HP1, while Buglia and Ferraro (2004) propose that it is the sperm that are tagged with proteins that determine sex. As an extension of intragenomic conflict between the sexes, haplodiploid fathers may favour

a female-biased sex ratio, as this will ensure the transmittance of their genes to offspring (de la Filia *et al.*, 2015). This may lead to male selection for the manipulation of PGE dynamics (de la Filia *et al.*, 2015). Contrastingly, females may favour a male-biased sex ratio due to resource competition between daughters (Taylor, 1981). Since female scale insects are sessile and longer-lived in contrast to short-lived mobile males, females require more resources for survival and reproduction (Taylor, 1981; Ross *et al.*, 2010a). Ross *et al.* (2010b), for example, showed that *Planococcus citri* (Coccoidea: Pseudococcidae) facultatively adjusted the sex ratio of offspring to favour males in response to high population densities.

Mutation rates and the efficiency of selection differentially affects haploid and diploid organisms (Otto and Gerstein, 2008). Relatively more mutations occur in diploids than haploids due to there being twice the number of genes available (Mable and Otto, 1998; Otto and Gerstein, 2008; de la Filia *et al.*, 2015). However, because of the capacity of homologous chromosome pairs to display heterozygosity, beneficial recessive mutations are masked and may take longer to become expressed in diploids (Mable and Otto, 1998). Additionally, the elimination of deleterious mutations through selection may be delayed due to this property (Otto and Gerstein, 2008). Because haploids immediately express mutations, deleterious mutations may be eliminated more efficiently, and beneficial mutations can spread more quickly through a haploid population (de la Filia *et al.*, 2015). Since haploid cells are generally smaller than diploid cells (Galitski *et al.*, 1999), haploids have a higher surface area to volume ratio. It has therefore been suggested that haploid organisms have the ability to take in more nutrients across their cell membranes and have a higher chance of survival under nutrient-limited conditions (Lewis, 1985). Additionally, a smaller genome size could be less metabolically demanding (Lewis, 1985). Cavalier-Smith (1978) suggested that smaller-bodied organisms, such as many endoparasites, tend to display haplodiploidy while larger-bodied organisms tend to be diploid due to differences in cell volume.

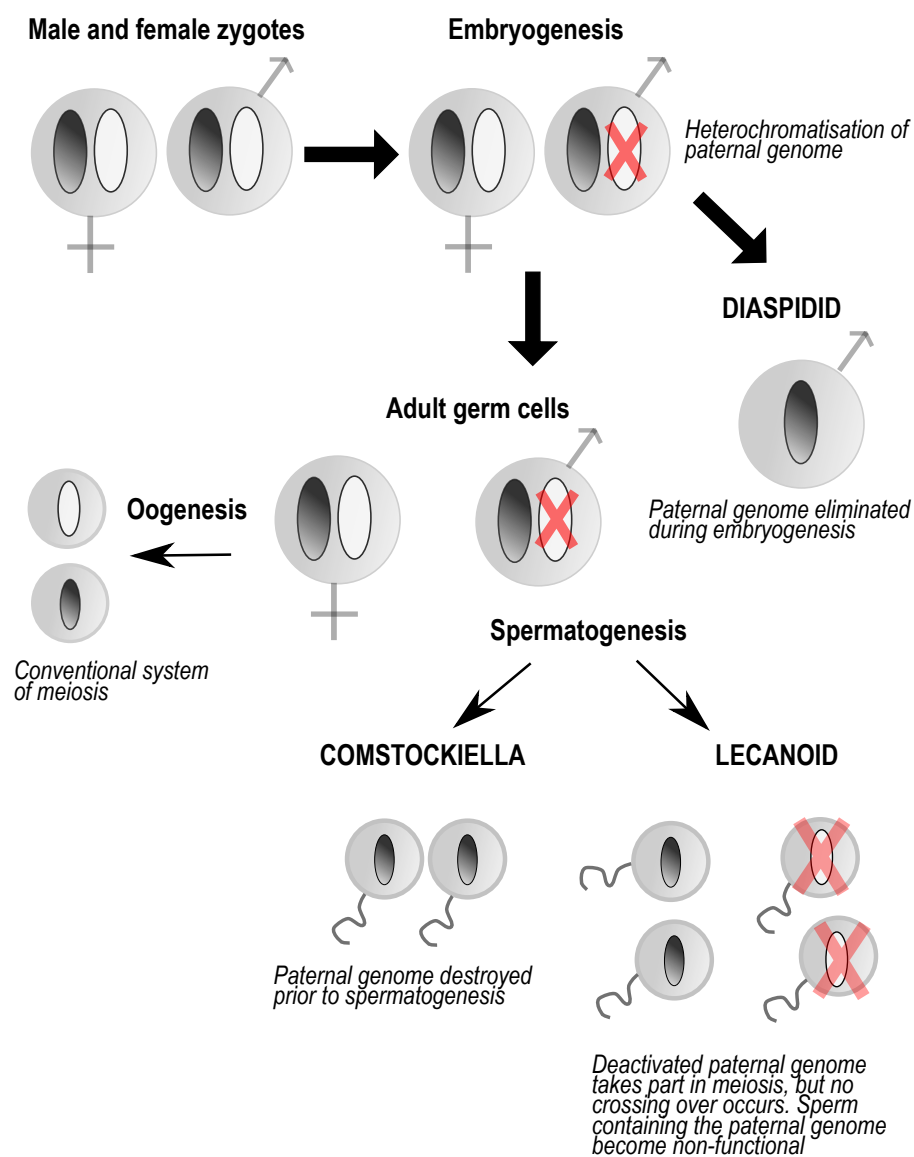
Scale insects feed on the phloem of their host plants, which is typically sugar-rich but nitrogen-poor (Buchner, 1965; Douglas, 2006). This has led to the evolution of obligate symbiotic relationships with microorganisms such as bacteria and unicellular fungi that synthesise amino acids and vitamins that their hosts can utilise (Buchner, 1965; Douglas, 1998, 2006, 2015). Endosymbiotic bacteria are usually housed in specialized organs called bacteriomes, in haemolymph, or inside modified fat cells within the insect hosts body (Buchner, 1965; Normark, 2004b; Rosenblueth *et al.*, 2018). Bacteria are vertically transmitted to offspring via the maternal line in the cytoplasm of eggs (Buchner, 1965; Rosenblueth *et al.*, 2018). Since males play no role in the

transferral of bacteriomes, it is theorized that a potential intergenomic conflict exists between male hosts and endosymbionts regarding sex allocation (Ross *et al.*, 2010b). Bacteria would favour a female-biased ratio to ensure their survival, while male insects would resist this pressure to prevent a reduction in paternal gene transfer to offspring (Ross *et al.*, 2010b). In PGE systems however, male hosts and endosymbionts may have a shared interest in a female-biased sex ratio as males can only transmit their genes to female offspring, as discussed above (Ross *et al.*, 2010b). Some studies have shown that symbionts can feminise (Rigaud, 1997) and even kill males (Hurst, 1991), or induce parthenogenesis in their host (Stouthamer *et al.*, 1990).

### 6.1.2 Paternal genome elimination

Paternal genome elimination involves the deactivation and elimination of the paternal half of the genome through the process of heterochromatisation in male offspring (Nur, 1980; Ross *et al.*, 2010b). As such, male sons transmit only their mother's genome to the next generation, making their closest male relatives their maternal grandfathers (de la Filia *et al.*, 2015). In this way, selection on male genes skips a generation (de la Filia *et al.*, 2015). At least 20 000 recorded species display PGE (de la Filia *et al.*, 2015), including more than half of the approximately 6000 known species of scale insects (Burt and Trivers, 2009). It has also been recorded in some fungus gnats (*Sciara* spp.) (Haig, 1993b), gall midges (*Mayetiola destructor*) (Haig, 1993a), phytoseiid mites (*Amblyseius* spp.) (Nelson-Rees *et al.*, 1980), springtails (*Sminthurus viridis* and *Allacma fusca*) (Collembola: Sminthuridae) (Dallai *et al.*, 2000) and a species of coffee borer beetle (Coleoptera: Scolytidae) (*Hypothenemus hampei*) (Brun *et al.*, 1995). There are three different categories of PGE, depending on the timing of the elimination of the paternal set of chromosomes in male offspring (Ross *et al.*, 2010b). These are namely the lecanoid, Comstockiella and diaspidid systems, where the Dactylopiidae family displays the Comstockiella PGE system (Nur, 1980) (Figure B2). The lecanoid and Comstockiella systems are very similar, the only difference is that the paternal half of the genome is eliminated during the process of spermatogenesis in the former, and before meiosis in the latter (Ross *et al.*, 2010b). In the lecanoid system, both the euchromatic maternal and heterochromatic (i.e. deactivated) paternal genomes undergo meiosis, but no crossing over occurs (Burt and Trivers, 2009). The resulting sperm cells that receive the paternal heterochromatic set are non-functional (Burt and Trivers, 2009). The Comstockiella system results in the full or partial elimination of the heterochromatic paternal genome prior to meiosis, where the number of paternal chromosomes that are eliminated can vary between species, individuals of the same species and even between

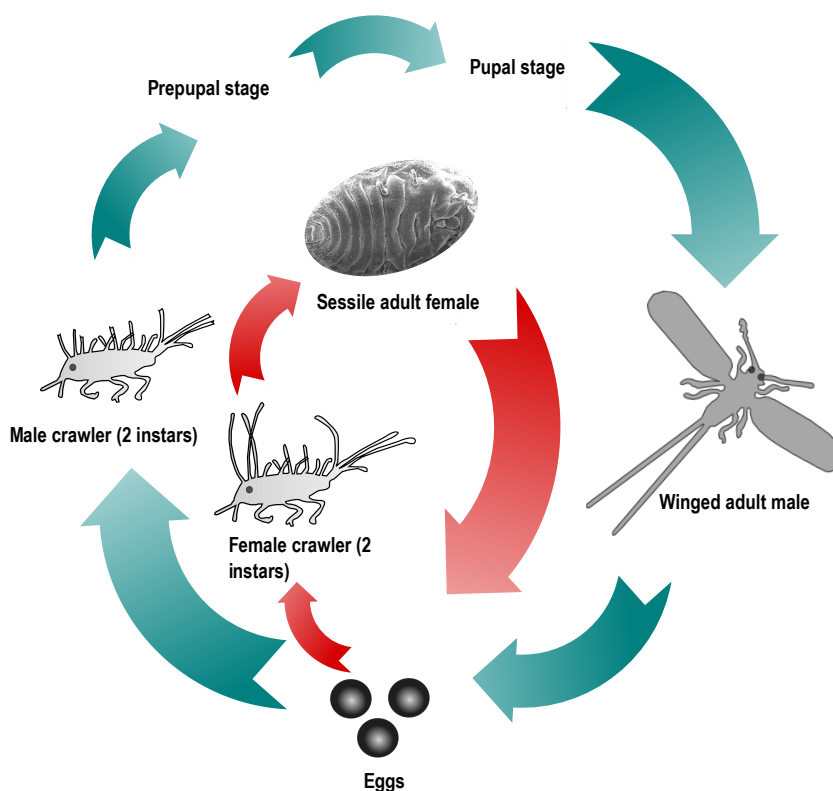
different regions of testicular cyst tissue within an individual (Burt and Trivers, 2009). The Diaspidid system has only been recorded in the Diaspididae family (the armored scale insects) (Nur, 1980; Ross *et al.*, 2010b; Burt and Trivers, 2009). In this system, instead of being deactivated, the paternal chromosomes are destroyed during embryogenesis, rendering males haploid throughout their lives (Ross *et al.*, 2010b). This process could have been selected for as an energy-saving strategy to eliminate the need of replicating deactivated chromosomes, and as a means of preventing their potential reactivation (Burt and Trivers, 2009). Nur (1967), for example, found that the male genome of the citrus mealybug *Planococcus citri* (Hemiptera: Pseudococcidae) became reactivated in some tissues, including reproductive cyst tissue where spermatogenesis takes place.



**Figure B2:** Paternal genome elimination in scale insects, showing the lecanoid, diaspidid and Comstockiella genetic systems. Adapted from Bull (1979); Burt and Trivers (2009); Ross *et al.* (2010b)

## 6.2 Life cycle of the Dactylopiidae

Pérez-Guerra and Kosztarab (1992) undertook a detailed study of the life history of *D. coccus*, which is representative of the genus. Males undergo six life stages (egg → first instar → second instar → pre-pupa → pupa → winged adult), while females undergo four (egg → first instar → second instar → sessile adult female). Oviposition usually occurs at night. Eggs are a light red colour and either hatch within the female (ovoviviparity), or within ten to thirty minutes after oviposition, yielding red nymphs (referred to as ‘crawlers’) approximately 1 mm in length (Nobel, 2002). Both male and female crawlers begin producing waxy filaments within an hour of hatching to assist in wind dispersal. Following attachment to a host plant, female nymphs produce a dense protective waxy layer after the second moult. Adult females lay an average of approximately 430 eggs in their lifetime, that they oviposit in chain-like formations on the surface of the host plant. Males display a holometabolous-like life cycle, which incorporates a pre-pupal and pupal stage, giving rise to winged adults (Nobel, 2002). Adult males do not feed, mate with multiple females, and live for approximately 50 to 80 days. Females are significantly longer-lived than males, living for up to five months (Moran and Cabby, 1979; Nobel, 2002).



**Figure B3:** Lifecycle of *Dactylopius*. Adapted from Moran and Cabby (1979)

## 6.3 Genetic barcoding application examples

**Table B1:** Cross-disciplinary examples of genetic barcoding

	Details
Medicine	Identification of the vertebrate hosts of haematophagous arthropods to better understand the ecology of host-parasite interactions and the implications for human and veterinary epidemiology (Alcaide <i>et al.</i> , 2009)
	Identification of sandfly (Diptera: Psychodidae) species in Columbia- vectors of Leishmania parasites (Gutiérrez <i>et al.</i> , 2014)
Agriculture	Barcoding of the Heliiothinae family to enable a method of rapid identification of potential crop pests entering Australia, particularly <i>Helicoverpa</i> Hardwick spp. and <i>Heliothis</i> Ochseneheimer spp (Ball and Armstrong, 2006)
	Identification of two species of <i>Busseola</i> (Lepidoptera: Noctuidae)- important pests of sugarcane in Ethiopia (Assefa <i>et al.</i> , 2007)
	Identification of invasive <i>Spodoptera</i> (Lepidoptera: Noctuidae) in North America (Nagoshi <i>et al.</i> , 2011)
Ecology	Barcoding of scats, hair, blood, tissue and regurgitates to determine the presence and dispersal of wolves recolonizing the western Alps as well as assigning haplotypes to different source populations (Valière <i>et al.</i> , 2003)
	Detecting the presence of a particular species of frog, <i>Rana catesbeiana</i> Shaw, in water samples taken from freshwater environments (Ficetola <i>et al.</i> , 2008)
	Determination of the diet of <i>Architeuthis</i> spp. The gut contents of a deep sea giant squid revealed the presence of <i>Macruronus novaezelandiae</i> Hector, and also indicated that cannibalism occurs in this squid genus (Deagle <i>et al.</i> , 2005)
	Analysing dirt DNA from soil samples as an indication of the vertebrate biodiversity in an area. The study concluded that the barcoding of eDNA can be an accurate measure of taxonomic richness and the relative biomass of species (Andersen <i>et al.</i> , 2012)
Forensic science	Detection of hairs from the protected Eurasian badger, <i>Meles meles</i> L., in shaving brushes produced in the Netherlands, France, Italy, the UK and Spain (Domingo-Roura <i>et al.</i> , 2006)
	Uncovering the mislabeling of seafood in restaurant outlets. Sushi restaurants sampled in Los Angeles were shown to have a mislabeling percentage of 47%, particularly regarding yellowfin tuna, yellowtail, Halibut and red snapper fish (Willette <i>et al.</i> , 2017)

Uncovering of illegal trade (by means of sampling sushi) in protected fin, sei and Antarctic minke whales between Japan and the USA and South Korea (Baker *et al.*, 2010)

---

Use of COI genetic barcoding to identify all the immature life stages of *Sarcophaga impatiens* Walker (Diptera: Sarcophagidae); an important insect in forensic entomology (Meiklejohn *et al.*, 2013)

---

Invasion biology

Use of COI and 16S barcoding regions to identify diapausing eggs in ballast water. 64% of the 96 samples were successfully identified to species level (Briski *et al.*, 2011)

---

Test of different plant barcoding genes to determine the highest hit rate for the identification of invasive plants in China. It was found that ITS and matK markers were the most effective (Xu *et al.*, 2017)

---

Paleontology  
and  
paleoecology

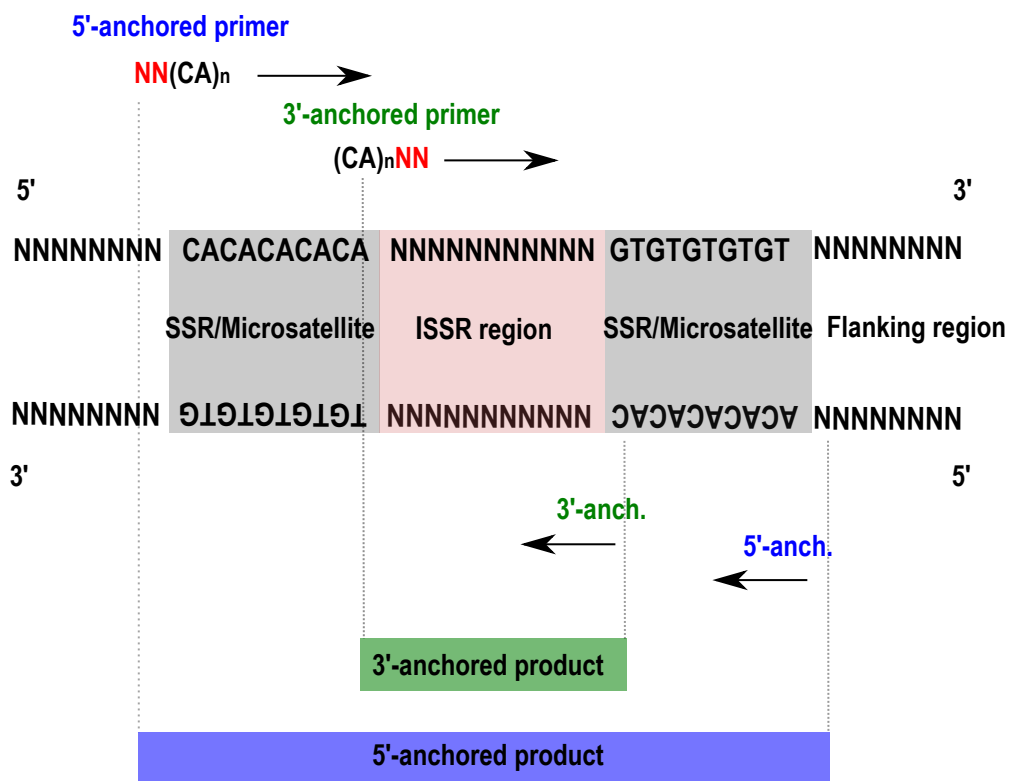
Analysis of the diet of Ötzi- the 5000 year old Neolithic mummy found in the Alps in 1991. The DNA barcoding results revealed that the man had eaten red deer (*Cervus elaphus* L.), ibex (*Capra ibex* L.) and cereals shortly before death. Pollen grains indicated that he must have passed through a subalpine forest during his journey (Rollo *et al.*, 2002)

---

Analysis of permafrost sediment cores from Siberia (dating from the present to ~1.5 to 2 million years ago), and cave and coastal sediments from New Zealand (dating back ~600 to 3000 years). Results revealed large scale changes in plant and faunal diversity and composition over time, as well as the presence of extinct megafauna such as ratite moas in New Zealand, and mammoths, bison and horses in Siberia (Willerslev *et al.*, 2003)

---

## 6.4 ISSR concept



**Figure B4:** An illustration of the concept of ISSR-PCR showing the use of 3' and 5'-anchored primers. A 5'-anchored primer attaches to the flanking region of a microsatellite, where a 3'-anchored primer attaches to the microsatellite region flanking the ISSR and forms a shorter PCR product. The ISSR region varies in size at different loci in the genome, producing unique banding patterns for each unique individual. Diagram adapted from Zietkiewicz *et al.* (1994); Reddy *et al.* (2002); Ng and Tan (2015).