Development of a molecular detection assay for accurate identification of five economically important tephritid species of commercial fruit in South Africa

By

Kelsey Jayne Andrews

Thesis presented in partial fulfilment of the requirements for the degree of Master of Science in the Faculty of Science at Stellenbosch University

Supervisor: Professor Hano Maree
Co-supervisor: Dr Rachelle Bester

December 2022

DECLARATION

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third-party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

December 2022

ABSTRACT

South African fresh fruit production and export are plagued by five major fruit fly pests (Diptera: Tephritidae). These fruit flies: *Ceratitis capitata* (Wiedemann) Mediterranean fly, *Ceratitis cosyra* (Walker) Marula fly, *Ceratitis quilicii* (de Meyer, Mwatawala & Virgilio) Cape fly, *Ceratitis rosa* (Karsch) Natal fly, and *Bactrocera dorsalis* (Hendel) the Oriental fly, are of quarantine significance in certain export markets. Apart from economic loss incurred due to limitations in export markets, these flies also cause physical damage to fresh fruit through oviposition. Accurate identification of these fruit flies can be tricky as this fruit fly family consists of multiple morphologically cryptic species and species complexes. Morphological identification of these fruit flies is unreliable when larvae or closely related species are intercepted. This has highlighted a need for accurate methods to distinguish between these species that do not rely on morphological traits. Therefore, this study aimed to develop a molecular identification assay that can differentiate between the five fruit flies of concern to South Africa accurately and timeously. To achieve this, two colony-reared insects from each species underwent DNA extraction and high throughput sequencing (HTS). HTS data were subjected to de novo assembly and used to construct ten complete mitochondrial genomes using a combination of de novo and reference-based assembly methods. From this, two identification assays were developed: a sequencing-based assay targeting a mitochondrial intergenic region and a multiplex PCR assay targeting the gene opsin Rh4. Regarding the sequencing-based assay, a single primer set was designed to amplify a mitochondrial region between $tRNA^{ile}$ and $tRNA^{met}$. The intergenic region between $tRNA^{ile}$ and $tRNA^{gln}$ (designated intergenic region I) within the amplicon is species-specific in size and proposed as a potential tool for species differentiation of the five species of interest in this study. In the multiplex PCR assay, five sets of species-specific primers with varying sizes were designed and optimised for use in a multiplex format. The resulting species-specific amplicons can be separated using a 2% agarose-TAE gel, providing accurate species identification. Both assays were validated using larval stages and wild, trap-collected specimens. The assays developed in this study can be applied in pest surveillance and monitoring activities and during fruit inspection at packhouses and ports of entry (PoE).

OPSOMMING

Suid-Afrikaanse varsvrugteproduksie en -uitvoer word geteister deur vyf groot vrugtevliegplae (Diptera: Tephritidae). Hierdie vrugtevlieë: *Ceratitis capitata* (Wiedemann) Mediterreense vlieg, *Ceratitis cosyra* (Walker) Marula vlieg, *Ceratitis quilicii* (de Meyer, Mwatawala & Virgilio) Kaapse vlieg, *Ceratitis rosa* (Karsch) Natalse vlieg, en *Bactrocera dorsalis* (Hendel) die Oosterse vlieg, is van kwarantynbelang in sekere uitvoermarkte. Afgesien van ekonomiese verliese wat gely word as gevolg van beperkings in uitvoermarkte, veroorsaak hierdie vlieë ook fisiese skade aan vars vrugte deur oviposisie. Akkurate identifikasie van hierdie vrugtevlieë is moeilik aangesien hierdie vrugtevliegfamilie uit veelvuldige morfologies kriptiese spesies en spesiekomplekse bestaan. Morfologiese identifikasie van hierdie vrugtevlieë is onbetroubaar wanneer larwes of naverwante spesies onderskep word. Dit het die behoefte aan akkurate metodes om te onderskei tussen hierdie spesies beklemtoon wat nie op morfologiese eienskappe staatmaak nie. Daarom het hierdie studie daarop gefokus om 'n molekulêre identifikasietoets te ontwikkel wat akkuraat en iiiining kan onderskei tussen die vyf vrugtevlieë wat vir Suid-Afrika kommerwekkend is. Om dit te bereik, het twee kolonie-insekte van elke spesie DNA-ekstraksie en hoë deurvloei-volgordebepaling ondergaan (HTS). HTS-data is aan de novo-samestelling onderwerp en gebruik om tien volledige mitochondriale genome te konstrueer deur 'n kombinasie van de novo- en verwysingsgebaseerde samestellingsmetodes te gebruik. Hieruit is twee identifikasietoetse ontwikkel: 'n volgordebepaling-gebaseerde toets wat 'n mitochondriale intergeniese gebied en 'n multipleks PCR-toets gerig op die geen opsin Rh4. Met betrekking tot die volgordebepaling-gebaseerde toets, is 'n enkele inleier-stel ontwerp om 'n mitochondriale gebied tussen $tRNA^{ile}$ en $tRNA^{met}$ te vermenigvuldig. Die intergeniese gebied tussen $tRNA^{ile}$ en $tRNA^{gln}$ (aangewese intergeniese streek I) binne die amplikon is spesie-spesifiek in grootte en kan as 'n potensiële hulpmiddel vir spesie-differensiasie van die vyf spesies van belang in hierdie studie. In die multipleks PCR-toets is vyf stelle spesiespesifieke inleiers met verskillende groottes ontwerp en geoptimiseer vir gebruik in 'n multipleks-formaat. Die gevolglike spesie-spesifieke amplikone kan geskei word met behulp van 'n 2% agarose-TAE-gel, wat akkurate spesie-identifikasie verskaf. Beide toetse is bekragtig deur gebruik te maak van larwes en wilde vlieë. Die toetse wat in hierdie studie ontwikkel is, kan toegepas word in plaagtoesig- en moniteringsaktiwiteite en tydens vruginspeksies by pakhuise en hawens of lughawens (PoE).

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I want to express my sincere gratitude and appreciation to the following people and institutions:

First and foremost, I am incredibly grateful to my supervisors, **Professor Hano Maree** and **Dr Rachelle Bester**, without whom this research would not have been possible. Thank you both for your excellent guidance, mentorship, and supervision. The skills, concepts, and opportunities you have provided me with are invaluable and have enabled me to become a capable and confident researcher.

Citrus research international for funding this project and their exceptionally welcoming environment and staff.

Dr Aruna Manrakhan for her expert knowledge of fruit flies and the provision of the specimens used throughout this study.

My fellow **lab mates and colleagues** in the Department, thank you for all your assistance and support and for making the working environment one of encouragement and positivity.

Lastly, to my incredible parents, **Nigel** and **Sandy** (who should soon be receiving golden globes for their supporting roles). The words “thank you” are simply insufficient for portraying my gratitude for all you have equipped me with. To my siblings, **Claudia**, **Ashleigh**, and **Graeme**, the last year has been unforeseeably challenging for us in many ways; thank you for conquering it by my side. And to **Daniel**, thank you for being my light.

LIST OF ABBREVIATIONS

µl	Microliter
µM	Micromolar
BLAST	Basic Local Alignment Search Tool
bp	Base pair
BRICS	Brazil, Russia, India, China, South Africa
COI	Cytochrome oxidase I
dNTP	Deoxynucleotide Triphosphate
DNA	Deoxynucleotide Nucleic Acid
EDTA	Ethylenediamine tetra-acetic acid
FARQ	Species complex (*Ceratitis fascientris*, *C. anonae*, *C. rosa* and *C. quilicii*)
HTS	High throughput sequencing
HRM	High-Resolution Melt
IGS	Intergenic spacer
IIT	Incompatible insect technique
IPM	Integrated pest management
IPPC	International Plant Protection Convention
mM	Millimolar
NCBI	National Centre for Biotechnology Information
ng	Nanogram
OTU	Operational taxonomic unit
PCG	Protein coding gene
PCR	Polymerase Chain Reaction
PoE	Port of Entry
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
SIT	Sterile insect technique
SNP	Single Nucleotide Polymorphism
TAE	Tris-acetic acid EDTA
Taq	*Thermus aquaticus* DNA polymerase
Tm	Melting temperature
tRNA	Transfer RNA
tRNA^{gln}	Transfer RNA ‘glutamine’
tRNA^{met}	Transfer RNA ‘methionine’

$tRNA^{ile}$	Transfer RNA 'isoleucine'
U	Unit
°C	Degree Celsius

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. INTRODUCTION

1.1. General introduction

1.1.1. Production and export of fresh fruit in South Africa

The climate in South Africa is subtropical and temperate, suitable for growing a variety of cereal crops, fruit, and vegetables. The South African fresh fruit industry is export-orientated, with the country being the second-largest citrus exporter worldwide (Citrus Growers' Association of Southern Africa, 2021). Fresh fruit trade, including both citrus and deciduous fruit, make up 35% of total agricultural exports from South Africa (Fresh Produce Exporters' Forum South Africa, 2022). BRICS countries, of which Brazil, Russia, India, China, and South Africa are members, represent the emerging economies of developing nations wherein agriculture is a central pillar of economic growth. In 2018, BRICS countries comprised 50% of the world's total agricultural output. Compared to the other four nations' agricultural output, South Africa holds advantage in the production and trade of fresh fruit as a result of its well-developed fruit and vegetable industries (Ren *et al.*, 2020). Adding on to this advantage South Africa also has the highest amount of agricultural land per capita, standing at 1.72 hectares per person. BRICS countries must ensure that their economies' agricultural and food sectors are sustainable. Sustainable agriculture is the optimisation of agricultural practices to decrease food and resource wastage (Kotze and Rose, 2015). The overwhelming majority of food waste occurs within the production, packaging, storage, and distribution phases (Von Bormann and Gulati, 2014). Phytosanitary pests such as tephritid fruit flies cause both direct and indirect losses. Direct damage is caused to fresh fruit during oviposition and larval feeding, in addition to premature fruit rot which occurs as a result of microorganisms entering oviposition sites (Badii *et al.*, 2015). Indirect losses occur through the restriction of export markets due to the risk of potentially introducing phytosanitary pests (Louzeiro *et al.*, 2021).

1.1.2. Barriers to trade

Phytosanitary and quarantine pests are defined by the International Plant Protection Convention (IPPC) as "any species, strain or biotype of plant, animal, or pathogenic agent injurious to plants or plant products". Such pests present in South Africa include citrus black spot (CBS) (*Guignardia citricarpa*) (Kiely), African citrus greening disease ('*Candidatus* Liberibacter africanus'), and certain mealybug species such as *Planococcus citri* (Risso). Internal fruit pests are arguably the most important quarantine pests as eggs, and larval stages can go undetected within the fruit itself: these include the false codling moth (FCM) (*Thaumatotibia leucotreta*) (Meyrick) and certain tephritid fruit fly species, *Ceratitis capitata* (Wiedemann), *Ceratitis rosa* (Karsch), and *Bactrocera dorsalis* (Hendel) (Grout, 2015). These quarantine pests are strictly monitored, and inspections are made to ensure that export produce is free from any pests that pose a phytosanitary risk to the importing country.

1.1.3. Practical implications of quarantine pests

Strict phytosanitary measures are often required for the export of fresh produce and are determined by the regulations of importing countries. Inspection, testing, and treatment of fruit are performed prior to export and again upon arrival at ports-of-entry (PoE) (ISPM 7, 2016). If the consignment meets all phytosanitary requirements, the produce is permitted entry into the country. However, failure to meet phytosanitary requirements and the detection of a quarantine pest are grounds for refusing entry. The consignment may then be detained at the port for further testing and inspection, reshipped at the exporting country's expense, or destroyed (incinerated) if it cannot be handled in another way. Unfortunately, current PoE identification technology is insufficient, and pests are commonly intercepted in a form that is not conducive to morphological identification, such as eggs, larvae, and damaged specimens (ISPM 20, 2019; Whatson, 2020). Where possible, early life stages are raised to adulthood for identification which may cause up to 14-day delays in the movement of fresh produce (ISPM 20, 2019). In cases where only genus-level identification is achieved, if a species within the genus is of quarantine significance the consignment is refused entry (Whatson, 2020). Continuous pest surveillance is essential in both exporting and importing countries to monitor population changes and determine pest-free areas. The pest status of a country continuously changes due to invasions or eradication, requiring amendments to existing phytosanitary import regulations.

1.2. Problem statement

There are five economically important fruit flies present in South Africa: *Bactrocera dorsalis*, *Ceratitis capitata*, *Ceratitis rosa*, *Ceratitis quilicii* and *Ceratitis cosyra,* which negatively affect all stages of production and export of fresh fruit from the country. To reduce the burden of fruit loss and losses in economic gain as a result of limiting export markets and shipment rejection, pest monitoring is performed to identify these flies and avoid further horticultural crop loss. Accurate pest identification is essential for monitoring population changes and making decisions regarding quarantine pests at PoE. However, current identification tools are unable to identify all five fruit flies, regardless of life stage, to species level simultaneously and timeously.

1.3. Aim and objectives

The aim of this study was to develop a molecular identification assay for assisting in the accurate identification of five fruit flies of economic significance to South Africa: *Ceratitis capitata*, *Ceratitis cosyra*, *Ceratitis rosa*, *Ceratitis quilicii* and *Bactrocera dorsalis*.

The general objectives of this study were as follows:

i. Collection of adult male specimens of each species from established colonies held at the Citrus Research International, Mbombela, South Africa.
ii. Generation of high throughput sequencing data for the five species of interest for the construction of a genetic database for each fruit fly.

iii. Development of a molecular assay for accurate identification and differentiation of the five fruit flies.

iv. Validation of molecular assay with wild, trap-collected specimens and larval stages.

1.4. Chapter overview

Chapter 1. Introduction

General introduction, problem statement and overview of the aims and objectives of the study are discussed. A chapter layout of the thesis is provided along, with the scientific outputs generated during this study.

Chapter 2. Literature Review

Firstly, a review of the agriculturally important fruit fly family Tephritidae, is undertaken including an overview of the *Bactrocera dorsalis* species complex and the *Ceratitis* FARQ species complex. This is followed by a timeline of existing research in molecular fruit fly identification as an evaluation of detection assays through the ages. Lastly, a look at existing molecular data and prospective areas for expanding current research.

Chapter 3. Utilisation of mitochondrial intergenic region for species differentiation of fruit flies (Diptera: Tephritidae) in South Africa

Mitochondrial genomes were assembled for all five target fruit flies. High-resolution melt analysis (HRM) and sequencing-based analyses were considered as species identification tools and validated using larvae and wild, trap-collected specimens.

Chapter 4. A multiplex PCR assay for the identification of fruit flies (Diptera: Tephritidae) of economic importance in South Africa

Whole genome assembly were performed, and target gene Opsin Rh4 was identified as a mechanism of species identification through multiplex PCR and validated using colony-reared larvae and wild, trap-collected specimens.

Chapter 5. Conclusion

Provides an overall summary of the main findings framed within the research context. The prospects for future research are discussed, along with final concluding remarks.

1.5. Research outputs

The following conference poster and research articles were generated during this study:

1.5.1. Conference poster

KJ Andrews, R Bester, A Manrakhan, HJ Maree. A new diagnostic tool for the identification of fruit fly larvae in citrus. Presented at the 11th Citrus Research Symposium, Champagne Sports Resort, 21-24 August 2022. Awarded best poster accolade. A copy of this poster is presented in Appendix C.

The work of chapters 3 and 4 contributed to this poster and was presented by KJ Andrews.

1.5.2. Publications

- Manuscript under review at BMC Genomics:

  Andrews KJ, Bester R, Manrakhan A, Maree HJ. Utilisation of mitochondrial intergenic region for species differentiation of fruit flies (Diptera: Tephritidae) in South Africa.

  This paper forms the basis of chapter 3. The first author (KJA) was responsible for the study's design, primer design, DNA extractions, HTS data analysis, mitogenome assembly and manuscript drafting. RB contributed to the study's design, primer design, HTS data analysis, mitogenome assembly and manuscript drafting. AM contributed to the study's design, supplied all colony and trap-collected fruit flies, and drafting of the manuscript. HJM contributed to the study's design and the manuscript's drafting.

- Article published in Scientific Reports:

  Andrews KJ, Bester R, Manrakhan A, Maree HJ. A multiplex PCR assay for the identification of fruit flies (Diptera: Tephritidae) of economic importance in South Africa. Scientific Reports. 2022;12:13089. A copy of the published article is presented in Appendix D.

  This paper forms the basis for chapter 4. The first author (KJA) contributed to the study's design, primer design for the multiplex PCR assay, DNA extractions, HTS data analysis, optimisation of the multiplex PCR assay, validation of the multiplex PCR assay, and drafting the manuscript. RB contributed to the study's design, primer design for the multiplex PCR assay, HTS data analysis, optimisation of the multiplex PCR assay, and drafting the manuscript. AM contributed to the study's design, supplied all colony and trap-collected fruit flies, and drafting of the manuscript. HJM contributed to the study's design and the manuscript's drafting.

1.6. References

Badii, K. B. *et al.* (2015) 'Review of the pest status, economic impact and management of fruit-infesting flies (Diptera: Tephritidae) in Africa', *African Journal of Agricultural Research*, 10(12), pp. 1488–1498. doi: 10.5897/AJAR2014.9278.

Von Bormann, T. and Gulati, M. (2014) *The Food Energy Water Nexus: Understanding South Africa's most urgent sustainability challenge*, *WWF-SA*. South Africa.

Citrus Growers' Association of Southern Africa (2021) '2021 Industry Statistics. 2020 Export Season'. Available at: www.cga.co.za.

Fresh Produce Exporters' Forum South Africa (2022) *Fresh Produce Export Directory*.

Grout, T. G. (2015) 'The Status of Citrus IPM in South Africa', *Acta Horticulturae*. Edited by B. Sabater-Muñoz, (1065), pp. 1091–1095. doi: 10.17660/ActaHortic.2015.1065.137.

ISPM 20 (2019) 'Guidelines for a phytosanitary import regulatory system', *IPPC, FAO*. Rome. Available at: www.ippc.int.

ISPM 7 (2016) 'Phytosanitary certification system', *IPPC, FAO*. Available at: www.ippc.int.

Kotze, I. and Rose, M. (eds) (2015) 'Reconnecting South Africa's food systems to its ecosystems', in *Farming Facts and Futures*. Cape TOWN, South Africa: WWF-SA.

Louzeiro, L. R. F. *et al.* (2021) 'Incidence of frugivorous flies (Tephritidae and Lonchaeidae), fruit losses and the dispersal of flies through the transportation of fresh fruit', *Journal of Asia-Pacific Entomology*, 24(1), pp. 50–60. doi: 10.1016/j.aspen.2020.11.006.

Ren, Y. *et al.* (2020) 'Development and Prospect of Food Security Cooperation in the BRICS Countries', *Sustainability*, 12(5).

Whatson, M. (2020) 'Decision To Revise Import Requirements for the Importation of Fresh Citrus From South Africa Into the United States', *Federal Registar*, 85(215).

CHAPTER 2. LITERATURE REVIEW

2.1. Tephritidae and the tainting of the fruit

True fruit flies (Diptera: Tephritidae) are an agriculturally important family, earning their name and ill repute from direct damage caused to plants and plant products while completing their life cycle. Roughly 30% of all tephritid flies are frugivorous attacking healthy fruit. Adult fruit flies feed by sucking fruit juice and honeydew (from aphids) off the surface of the fruit without puncturing it (Christenson and Foote, 1960). Adult female flies have a large, elongated ovipositor which they use to pierce the fruit skin and deposit their eggs in clutches. Once the eggs have hatched, larvae tunnel towards the centre of the fruit while feeding on the soft pulp. During this stage, the fruit is vulnerable to secondary infections causing it to decompose and drop to the ground. The larvae undergo three instar stages before they emerge from the fruit by dropping into the soil to pupate. Adult flies emerge from pupae on roughly day 36 of their life cycle depending on temperature (Ekesi and Billah, 2006) (Fig. 2.1). Females frequently make multiple probes before selecting the final oviposition site and these small cavities are often exploited by secondary pests and fungal or bacterial infections which cause further fruit rot (Badii *et al.*, 2015). Bacteria from the intestinal gut flora of the female fly are introduced to the fruit during oviposition, and the resulting decomposition makes the fruit pulp soft for the developing larvae to feed. External signs of fruit fly damage are marked by the browning of the surface of fruit around oviposition punctures and soft spots on the fruit caused by internal degradation (Manrakhan, 2020). Visible signs of damage to commercial fruit significantly reduce the market value. Tephritid flies are most commonly intercepted in immature life stages, such as larvae and eggs (Whatson, 2020). These life stages are easily overlooked at packhouses and during inspections as the early life stages are concealed within the fruit. The developmental time of tephritid flies is entirely dependent on favourable climatic conditions; during the colder winter months developmental time is considerably slower. Such that cold treatment and cold storage of commercial fruit, primarily used for preserving fruit quality, are carried out at low temperatures (>3°C) in attempts to halt fruit fly development (Grout *et al.*, 2011; Follett and Snook, 2013).

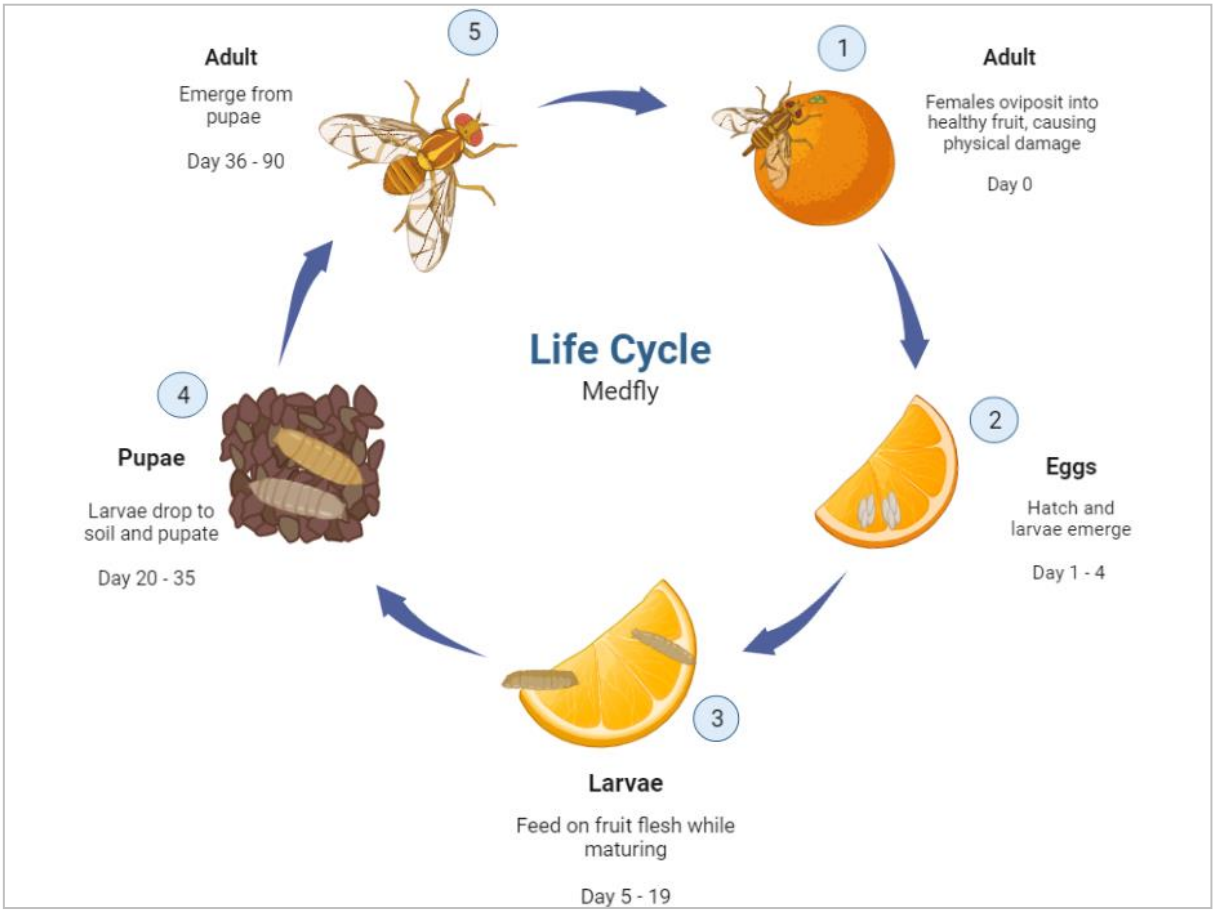


Fig. 2.1. Diagrammatic representation of the life cycle of a tephritid fruit fly. The approximate developmental time for each major life stage is displayed for summer months, development times during winter months are considerably longer. Image was created with BioRender online software.

2.2. The complex case of Tephritidae

Historically, morphology formed the basis for the taxonomic delineation of species (Lukhtanov, 2019). This can be problematic when boundaries between species are ambiguous. The availability and applicability of molecular data used for the phylogenetic reconstruction of taxonomic trees have, in many cases, resulted in both the synonymising and splitting of species (Schutze *et al.*, 2015; De Meyer *et al.*, 2016). Species complexes can refer to groups of morphologically indistinguishable cryptic species or sibling complexes of closely related species (Schutze *et al.*, 2017). The presence of species complexes often confounds species-level resolution in phylogenetic studies as these morphologically alike, closely related species fail to resolve as separate clades (Virgilio *et al.*, 2008). The presence of species complexes and cryptic species hinder the success of fruit fly-specific control strategies for morphologically similar species with different host ranges and ecological requirements.

2.2.1. The *Bactrocera dorsalis* species complex

The *Bactrocera dorsalis* species complex is of Asian origin and consists of close to 100 morphologically similar species (Drew and Hancock, 1994). Of the vast number of members within this complex, only five are of significant economic importance due to damage caused to commercial fruit and vegetables; they are

Bactrocera dorsalis (Hendel), *Bactrocera papayae* Drew & Hancock, *Bactrocera philippinensis* Drew & Hancock, *Bactrocera carambolae* Drew & Hancock, and *Bactrocera invadens* Drew, Tsuruta & White (Drew and Romig, 2013). Following thorough examination, *B. papayae*, *B. philipinensis* and *B. invadens* have been synonymised with *B. dorsalis* due to evidence suggesting the variation observed between these former species is not species-level variation but rather population-level variation (Schutze *et al.*, 2015). Morphological differences between these four synonymised species are not consistent and are variable within species. Furthermore, these four species cannot be resolved using DNA barcoding, and fertile hybrids are produced from random mating between these flies showing no reproductive isolation (Schutze *et al.*, 2015). The large amount of intraspecific variation observed in colour patterns in species of the *Bactrocera dorsalis* species complex restricts the utility of morphological keys (Leblanc *et al.*, 2015). This is concerning as morphology is used as the basis of species identification for pest monitoring and decisions of quarantine importance.

2.2.2. The history of the *Ceratitis* FARQ complex

Today, the *Ceratitis* FARQ complex is a small group of four fruit flies of Afrotropical origin, *Ceratitis fasciventris* (Bezzi), *Ceratitis anonae* Graham, *Ceratitis rosa* (Karsch) and *Ceratitis quilicii* de Meyer, Mwatawala & Virgilio. In the past, before the separation of the four species, *C. fasciventris*, *C. quilicii* and *C. rosa* were classified as a single species, *C. rosa* s.l. In 1920, *C. fasciventris* was described as a variant of *C. rosa* s.l, (*C. rosa* var. *fasciventris* (Bezzi)) (Bezzi, 1920). It was only in 2001 that *C. fasciventris* was described as a separate species (De Meyer, 2001). The *Ceratitis* FAR complex was formed (*C. fasciventris*, *C. anonae* and *C. rosa*). These species are morphologically very similar, with minor differences present in the colouration and feathering patterns of the mid tibia of male specimens (Fig. 2.2.). *Ceratitis rosa* was further split into two populations, R1 and R2, based on allelic variation at 16 microsatellite markers (Virgilio *et al.*, 2013). The R1 population was found to be more abundant at low altitudes, and the R2 type was more abundant at higher altitudes (Mwatawala *et al.*, 2015). In 2016 the distinction of two separate species was made, *C. rosa* (formerly R1, the lowland type) and *C. quilicii* (formerly R2, the highland type) (De Meyer *et al.*, 2016). Within the FARQ, complex only adult male specimens are morphologically identifiable. Larval stages and female flies are near identical, with the exception being *C. anonae*. *Ceratitis rosa* is described as a pest of quarantine significance. Due to the recent separation of *C. rosa* and *C. quilicii*, the latter fly is afforded the same quarantine status as its sibling species until further information is gained about its host status and distribution (Virgilio *et al.*, 2013).

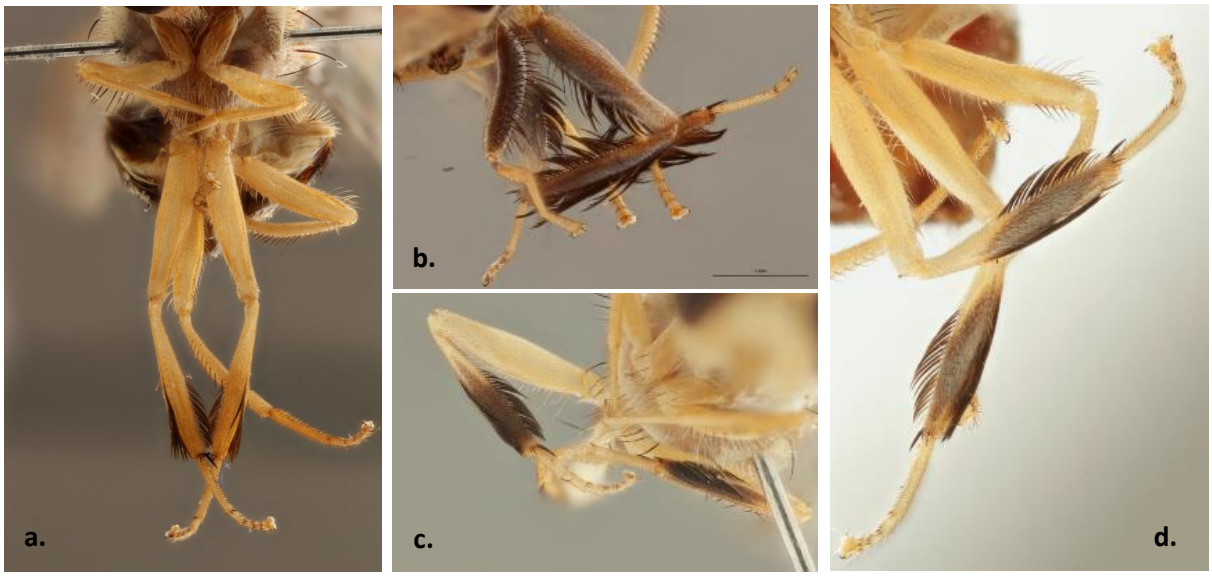


Fig. 2.2. Visual demonstration of similarity in morphological traits used for species identification of the *Ceratitis* FARQ complex. The following figures are an anterior view of the legs (focused on the mid tibia) of male specimens of each species. **a.** *Ceratitis fasciventis*. **b.** *Ceratitis anonae*. **c.** *Ceratitis quilicii* mid tibia is more slender with black colouration not reaching ventral and dorsal margins (De Meyer et al., 2016). **d.** *Ceratitis rosa* mid tibia is broad with black colouration reaching ventral and dorsal margins. Images adapted from the Royal Museum for Central Africa online collection. ©Africa Museum, image author: Johnathan Brecko.

Interrogations into whether this species complex truly consists of separate species or whether population level variation has been over split are ongoing. Under laboratory conditions, *C. fasciventris* and *C. rosa* s.l (experiments performed prior to 2016) were able to form fertile hybrid offspring (Erbout, De Meyer and Lens, 2008). Although no evidence of hybridisation of these two species has been observed in the field, they do share a large overlap in host range and have frequently been reared from the same host fruit (Copeland *et al.*, 2006). More recent genetic studies have shown a lack of reproductive isolation between members of the FARQ complex illustrated by gene flow estimates between species, suggesting potential admixture events, particularly between *C. fasciventris* and *C. quilicii* (Virgilio *et al.*, 2013; Zhang *et al.*, 2021). These findings, combined with the fact that FARQ fruit flies share largely overlapping ecological distributions, bring into question the ecological and biological species concept. The phylogenetic species concept of this complex has recently been reaffirmed by Zhang *et al.* (2021), where monophyly of each species was resolved using genome-wide single nucleotide polymorphism (SNP) data (Fig. 2.3.).

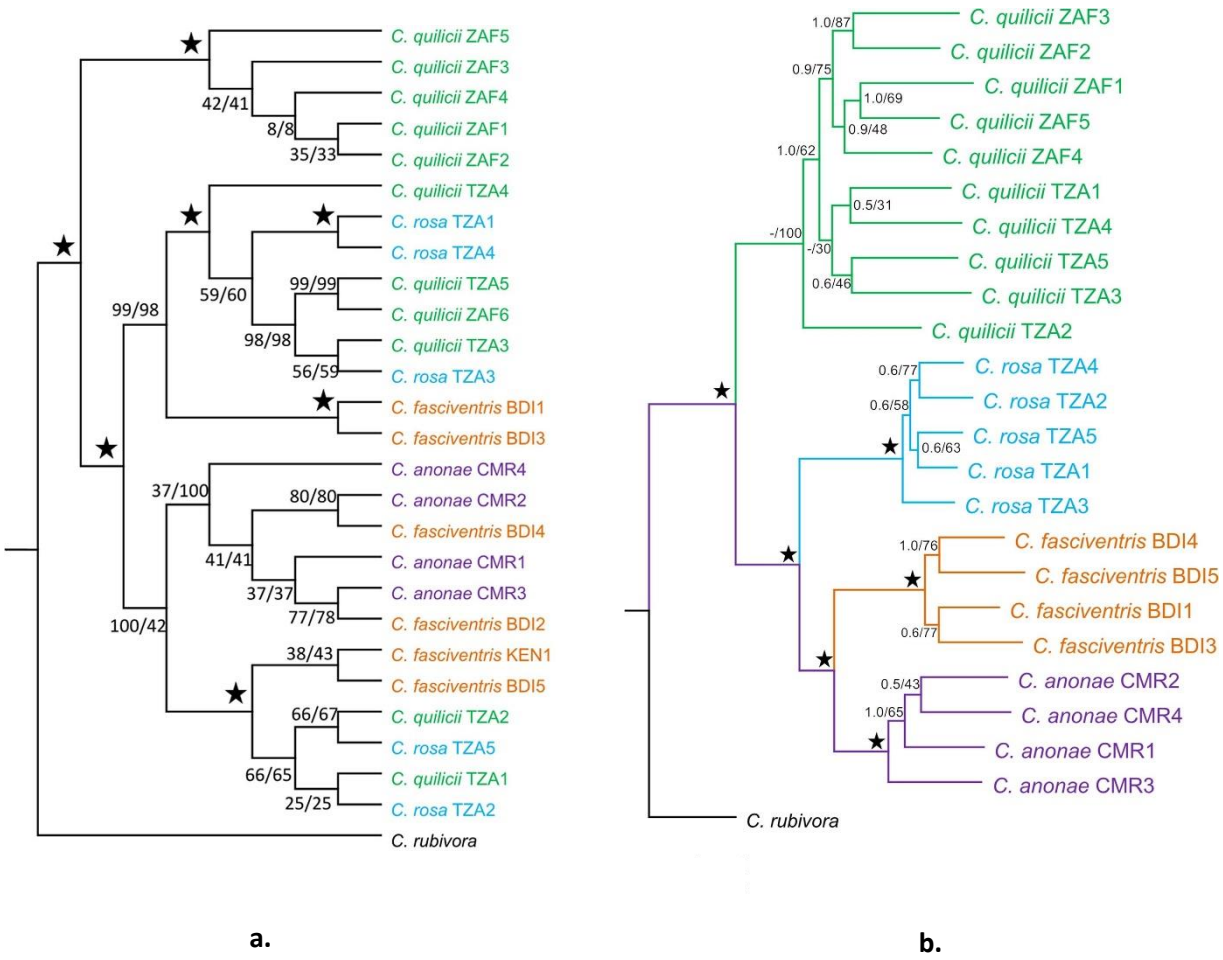


Fig. 2.3. Phylogenetic resolution of the FARQ complex. **a.** Phylogenetic analyses performed using mitochondrial protein-coding genes (dataset included all 13 PCG's and two rRNA's) was unable to resolve the FARQ complex into species-specific monophyletic clades. **b.** Phylogenetic analyses performed using genome-wide SNP data allowed greater phylogenetic resolution of the FARQ complex into distinct monophyletic clades. Phylogenetic trees are adapted from Zhang *et al.* (2021).

2.2.3. Species complexes, endosymbionts, and speciation

Wolbachia are endosymbiotic bacteria known for causing cytoplasmic incompatibility (CI) in hosts. This bacterium is primarily spread through vertical transmission down the maternal line. Mating between males infected with *Wolbachia* and uninfected females, or females infected with a different strain of *Wolbachia*, result in the production of non-viable embryos causing mating disruption (Weeks, Reynolds and Hoffmann, 2002). It is suggested that Wolbachia could be a driving force in the diversification of species complexes, as was demonstrated with the *Anastrepha fraterculus* (Wiedemann) species complex in South America. In this example, morphotypes of the *Anastrepha fraterculus* species complex were shown to harbour different strains of *Wolbachia* (Prezotto *et al.*, 2017). Similarly the *Anastrepha fraterculus* complex and the *Ceratitis* FARQ complex have both exhibited mating incompatibility between morphotypes, where hybrid progeny have low viability (Erbout, De Meyer and Lens, 2008; Virgilio *et al.*, 2013; Prezotto *et al.*, 2017). The zygotic incompatibility seen between crosses of these morphospecies may be an indication of the effects of *Wolbachia*. Furthermore, the embryotic mortality phenomenon has been exploited and trialled as a mechanism of pest control known as the incompatible insect technique (IIT), where mass-reared males

infected with a certain strain of *Wolbachia* are released to mate with uninfected females or females infected with a different *Wolbachia* strain (Boller *et al.*, 1976; Mateos *et al.*, 2020). In addition to mating disruption, population dynamics are also affected by geographic barriers which prevent certain *Wolbachia* strains from migrating and infecting isolated populations, as was studied in *Rhagoletis cerasi* Loew in Europe (Riegler and Stauffer, 2008). The prevention of gene flow between populations through geographical isolation, combined with Wolbachia induced mating disruption may be a potential avenue of speciation and cause of divergence which may result in the presence of cryptic species and species complexes (Keeling, Jiggins and Read, 2003). Mechanisms of potential horizontal transmission of *Wolbachia* include the hybridisation of closely related species, sharing of ecological niches and food sources, and sharing of predators and parasites (Morrow *et al.*, 2014; Bruzzese *et al.*, 2022). By extrapolation, it is plausible that closely related species sharing a similar ecological niche should have more strains of *Wolbachia* in common than species which are not closely related and do not share overlapping ecological niches. Hence, it may be possible to identify host specimens, including cryptic species and species complexes, to species level based on the diversity of *Wolbachia* strains present in their microbiome.

2.3. Fruit fly pests of South Africa

2.3.1. Pests of non-priority export fruits

Minor tephritid pests are those which don't make use of commercial produce as host plants, have a limited host range, or are not tolerant of differing climatic conditions. These fruit flies make up the majority and are typically not of quarantine significance, and do not necessitate phytosanitary action. *Ceratitis quinaria* (Bezzi) affects a limited range of fresh fruit and is contained within the African continent. *Dacus ciliatus* Loew is a pest of cucurbits (White and Elson-Harris, 1992), although this fly shows potential to establish itself outside of its native range, cucurbits are not one of the major export commodities of fresh fruit in South Africa. *Bactrocera oleae* (Gmelin) is a monophagous pest of olives and has established itself in all main olive-producing regions. The movement of this pest is associated with the introduction and transport of olive trees rather than the fruit itself (Nardi *et al.*, 2005).

2.3.2. Pests of major export fruits

The major pests affecting the production and export of fresh fruit from South Africa comprise five frugivorous tephritid flies, *Ceratitis capitata* (Wiedemann), *Ceratitis cosyra* (Walker), *Ceratitis rosa*, *Ceratitis quilicii* and *Bactrocera dorsalis* (Hendel) (Fig. 2.4.). The Mediterranean fruit fly, *C. capitata*, is highly polyphagous and invasive attacking a wide variety of unrelated commercial fruit such as citrus, mango, apples, pears, and peaches (White and Elson-Harris, 1992; De Meyer *et al.*, 2002). It is the most serious fruit fly pest, able to tolerate harsh climatic conditions and is widespread throughout Africa, South America, Western Australia, and Europe (EPPO, 2022). The marula fly, *C. cosyr*a, is one of the most destructive pests of mango and currently has a limited Afrotropical distribution. This fly has frequently been intercepted at ports of entry (PoE) (Whatson, 2020) and is of invasive concern for China (Li *et al.*, 2009). The Natal fly *C. rosa*, like *C.*

capitata, is highly polyphagous, attacking a similar range of unrelated commercial fruit, and in some instances, it has displaced *C. capitata* from regions in which they both occur (Duyck, David and Quilici, 2004). The Natal fly is currently restricted to Southern Africa, with a high potential for invasiveness (EPPO, 2022). The Cape fruit fly, *C. quilicii*, has recently been described, and its host range is still being determined. However, this fly has already exhibited expansion to Mauritius and Reunion islands (Tanga *et al.*, 2018). The oriental fruit fly, *B. dorsalis*, is a highly invasive species, spread throughout sub-Saharan Africa, Asia, and Melanesia (EPPO,2022), posing a significant risk for expansion to Europe. *Bactrocera dorsalis* is exotic and established itself in South Africa in 2013 (Manrakhan, Venter and Hattingh, 2015). This is another greatly polyphagous species which causes significant damage to mango, guava, and citrus.

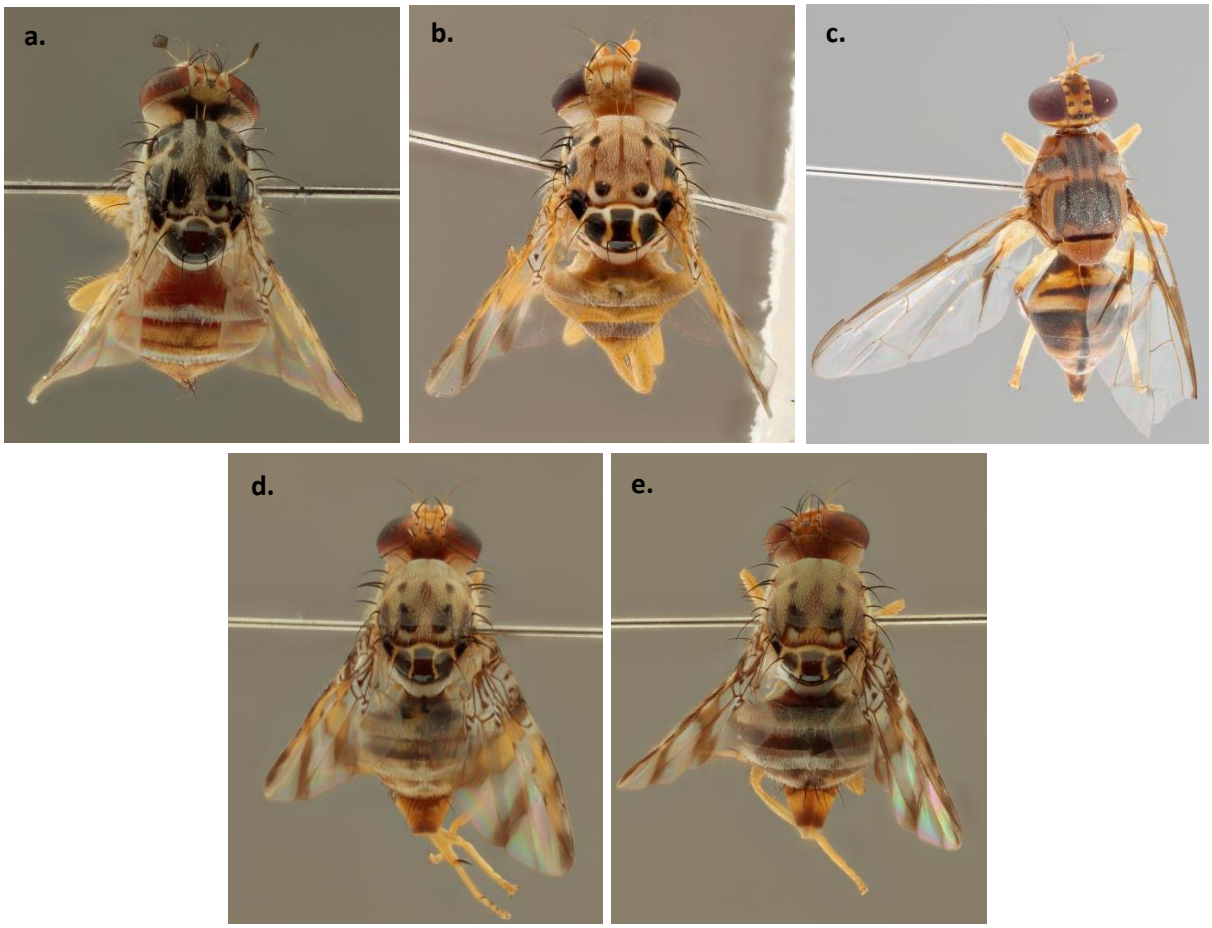


Fig. 2.4. Visual display of five fruit flies of major economic importance to South Africa. **a.** *Ceratitis capitata*. **b.** *Ceratitis cosyra*. **c.** *Bactrocera dorsalis*. **d.** *Ceratitis rosa*. **e.** *Ceratitis quilicii*. Note: *C. rosa* and *C. quilicii* are cryptic species which are morphologically identical apart from colouration of the midtibia in male specimens (Fig. 2.2.). Images adapted from the Royal Museum for Central Africa online collection. ©Africa Museum, image author: Johnathan Brecko.

2.4. Fruit fly management

2.4.1. Integrated pest management (IPM)

Integrated pest management (IPM) is the use of multiple practices in combination for the most effective action against pests. Examples of IPM practices include:

- Fruit fly monitoring using baited traps or attractants, helps to determine which species are present in an area and provides an estimate of population size.
- Maintaining sanitation in orchards and vineyards by collecting and removing fallen or damaged fruit to prevent attracting fruit flies to the area.
- Mechanical protection, such as wrapping or bagging fruit prevents oviposition, as fruit flies cannot directly access the fruit.
- Harvesting at a stage of maturity of fruit that isn't optimal for oviposition is another mechanism proven effective with *B. dorsalis* and harvesting unsuceptible green bananas (Cugala *et al.*, 2014).
- Chemical control via spraying canopies with bait spray as a deterrent for flies and other pests (Manrakhan, 2020).
- Biological control through the release of known parasitoids and pathogens of tephritid fruit flies to reduce fruit fly populations, this mechanism is not practical at commercial scale in South Africa.
- Mating disruption via sterile insect technique (SIT) and pheromone baited traps to reduce insect population sizes.
- Cold-treatment and cold storage post-harvest make conditions unfavourable for larval development (Grout *et al.*, 2011; Follett and Snook, 2013).

2.4.2. Eradication and pest-free zones

Declaring pest-free zones has advantages for trade through reducing the risk of introducing potentially invasive flies (ISPM 35, 2018). The distribution of the five fruit flies of economic concern in South Africa largely overlap, with *C. capitata* being widespread throughout the country (De Villiers *et al.*, 2013), *C. cosyra* is restricted to the Northeast and Eastern regions of the country, *C. rosa* and *C. quilicii* largely overlap in distribution with *C. quilicii* being dominant in cooler regions and at higher altitudes (De Meyer *et al.*, 2016). *Bactrocera dorsalis* has been effectively eradicated from parts of South Africa, including the Addo area in the Eastern Cape, the Grabouw region of the Western Cape, and along the Orange River in the Northern Cape (www.ippc.int). This was achieved through the use of orchard sanitation, bait sprays, and male annihilation (Manrakhan, Venter and Hattingh, 2015). The suppression of fruit fly populations can also be achieved with the sterile insect technique (SIT), under which artificially sterilised males are released into the wild to mate with fertile females to reduce the number of potential offspring produced. SIT is currently underway with *C. capitata* to establish the first fruit fly-free zones in South Africa (Barnes and Venter, 2006; Grout, 2015). Continuous monitoring is necessary to determine the success of these management programs, the distribution of these fruit flies, and determine the potential entry of other exotic flies to the country through the exotic fruit fly surveillance program (Barnes and Venter, 2006).

2.5. The morphological identification quandary

The identification of tephritid fruit flies to species level is generally done using taxonomic keys. Such keys have been developed for identifying both adult and larval life stages (Virgilio, White and De Meyer, 2014;

Balmès and Mouttet, 2017). However, even multi-entry keys, where a user can decide to either skip characters or submit multiple answers on uncertain characteristics, do not guarantee unambiguous identification of species. For instance, females of the subgenus *Ceratitis (Pterandrus)*, which includes the *Ceratitis* FARQ complex, are morphologically cryptic and cannot be identified using structural characteristics (De Meyer and Freidberg, 2005; Virgilio, White and De Meyer, 2014). Conversely, larval identification keys are only reliable when used by entomologists who specialise in Tephritidae and cannot distinguish *C. cosyra* from the members of the FARQ complex (Balmès and Mouttet, 2017). The limitations for the sole use of morphology to identify tephritid flies to species level are apparent, specifically where phenotypic, morphological variation in characteristics predominantly used for identification can result in misidentifications and over-splitting or over-lumping of genetically distinct species (Funk and Omland, 2003; Hebert *et al.*, 2003). The over-splitting of species occurs when intraspecific or population-level variation is mistaken for species-level variation, as occurred with certain members of the *Bactrocera dorsalis* complex (Tan *et al.*, 2013). The over-lumping of species is the inverse, where species-level variation is mistaken for intraspecific or population-level variation, as was the case with the *C. rosa* s.l, before being split into *C. rosa*, *C. fasciventris* and eventually *C. quilicii*. It is evident that in instances where morphological differences between species are slim, female specimens are collected, specimens are damaged or missing characteristic parts, or early life stages are found, the use of molecular tools can aid existing morphological species diagnostics to improve accuracy and avoid erroneous classifications.

2.6. Molecular identification tools: a brief history

Molecular-based identification tools have begun replacing the need to rear intercepted larval specimens to adult life stages, which is a largely unsuccessful and time-consuming process for making time-sensitive biosecurity and quarantine decisions (Armstrong and Ball, 2005). The use of molecular tools can save up to two weeks on diagnostic requiring decisions, saving not only time but the quality and shelf-life of export produce. Below a timeline of advances in tephritid molecular identification is displayed (Fig. 2.5.).

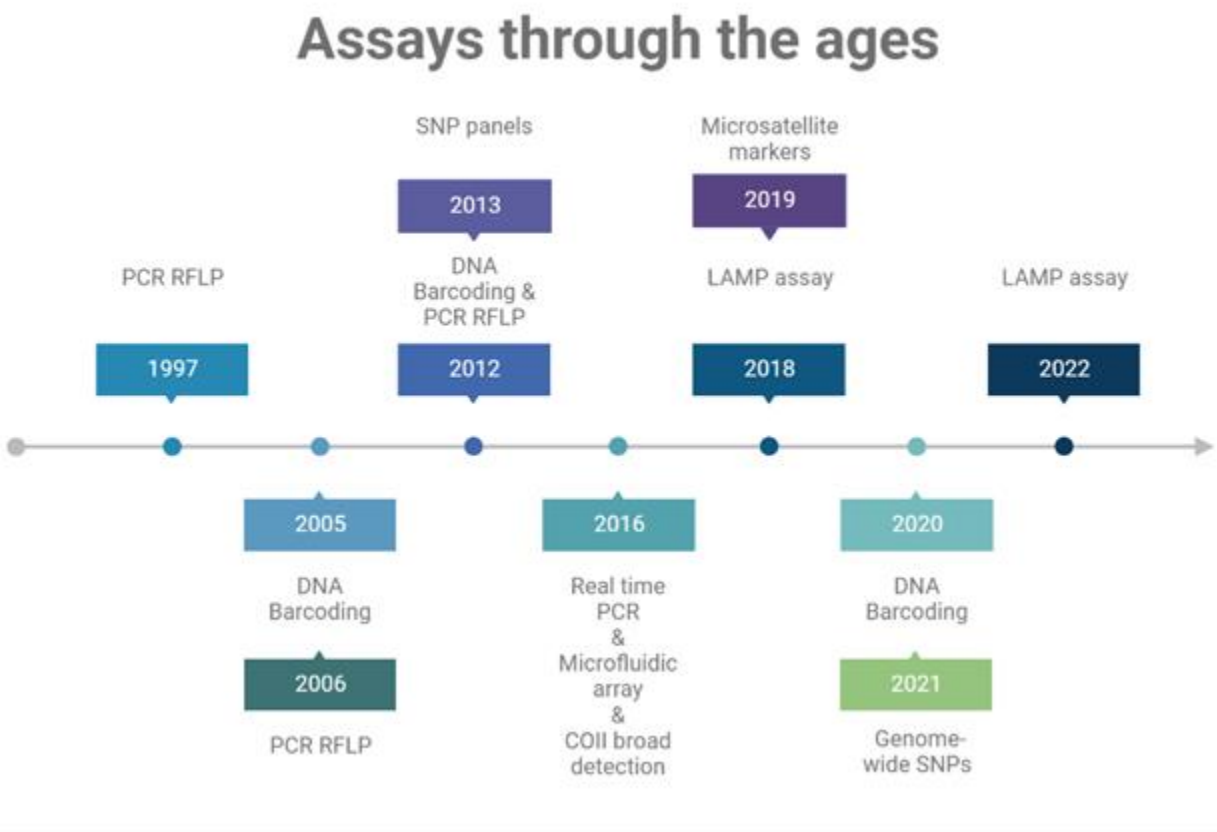


Fig. 2.5. Representative timeline of molecular tools for fruit fly identification most relevant to the context of the current study: an overview of assays through the ages.

2.6.1. PCR-RFLP

Restriction fragment length polymorphism (RFLP) refers to a technique in which DNA samples are digested into fragments using selected restriction enzymes. Differences or polymorphisms between DNA samples at restriction fragment recognition sites will result in fragment length differences between the DNA samples. The first PCR-RFLP reported for tephritid flies was developed as a rapid identification tool for detecting immature life stages of a group of economically important flies in New Zealand. This assay targeted the 18S and 18S+ internal transcribed spacers (Armstrong, Cameron and Frampton, 1997). The restriction analysis resolution of the 18S was poor and could not be used as a diagnostic for species differentiation. However, the digestion of the 18S+ internal transcribed spacer was able to differentiate between 13 of the 19 species of interest in the study. Later, PCR-RFLP was employed for the analysis of mitochondrial genes, 12S rRNA, 16S rRNA, and NAD6, to differentiate between *Ceratitis* species (Barr *et al.*, 2006). This tool was able to routinely differentiate between most species surveyed except for *C. capitata* and *C. caetrata* Munro and the FAR complex (*C. fasciventris*, *C. anonae* and *C. rosa*). The efficacy of PCR-RFLP was then compared against DNA barcoding for determining the best resolution of the genus *Ceratitis* (Barr *et al.*, 2012). Although both methods performed similarly, PCR-PFLP was more economically appropriate due to extra sequencing and reagent costs encountered with DNA barcoding.

2.6.2. Cytochrome oxidase I based detection

2.6.2.1. DNA barcoding

DNA barcoding is the amplification of the complete or partial segment of cytochrome oxidase I (COI) to determine the identification of a sequence of interest by querying it against a database of previously identified COI sequences. This comparison is made using genetic distance; hence a query sequence and best match will share a high similarity and ideally, a low genetic distance. Early DNA barcoding in Tephritidae began with *Bactrocera*, which was found to be limited in its ability to differentiate between species within species complexes (Armstrong and Ball, 2005; Boykin *et al.*, 2012; Jiang *et al.*, 2014). Studies regarding the genus *Ceratitis* had similar results with an inability to differentiate between *C. capitata* and *C. caetrata*, and between the members of the FARQ complex (De Meyer *et al.*, 2016; Balmès and Mouttet, 2017). However, the accuracy of DNA barcoding is dependent on the reliability of the sequence similarity approach. High throughput sequencing technology (Pacific Bio SMRT sequencing) has been compared to Sanger sequencing for DNA barcoding utility, whereby primers designed to amplify COI were modified with a 30 bp universal tag to be compatible with PacBio's circular consensus sequencing (Garzón-Orduña *et al.*, 2020). The amplicons generated with Sanger sequencing were compared against those generated with HTS. Genus level identification was congruent between the two methods 99% of the time, while species level identification was congruent 98% of the time. The limitations of DNA barcoding with COI, are likely due to unclear species boundaries between closely related species and to the existence of cryptic species in Tephritidae (Jiang *et al.*, 2014). Low success rates of DNA barcoding with species complexes are attributed to overlapping inter and intraspecific genetic distances. In fact, DNA barcoding has been suggested to potentially identify cryptic species in cases where there is high intraspecific variation within COI (Waugh, 2007). *Ceratitis cosyra* is suspected of harbouring cryptic speciation as DNA barcoding results of this species split the species into two different monophyletic groups (Virgilio *et al.*, 2017). Apart from this, COI does perform well in distinguishing between species which do not exist in complexes.

False positive results can easily occur in cases where a query specimen is not represented by an existing DNA barcode in the database, and a false positive result is assigned based on the highest sequence identity leading to erroneous classification. As DNA barcoding is susceptible to both false-positive and false-negative results (Boykin *et al.*, 2012), the use of a distance threshold to minimise false-positive results has been suggested, where if the distance between a query sequence and best match exceeds the threshold value set, the identity is considered unreliable and is discarded. However, the improvement in precision comes at the sacrifice of potentially valuable data as a result of large numbers of queries discarded (Virgilio *et al.*, 2012). Overall, DNA barcoding is a robust technique that can be applied across a wide range of organisms. However, it is a time-consuming process, from the interception of the larvae to the identification of species (where possible). Unfortunately, many quarantine laboratories do not have sequencing facilities on site (Boykin *et al.*, 2012), adding additional expenditure and waiting time. As applicable as DNA barcoding is in identifying some

species, COI is not diverse enough to permit unambiguous identification of all tephritid flies (Blacket, Semeraro and Malipatil, 2012).

2.6.2.2. SNP Panels

Single nucleotide polymorphisms or SNPs are genetic variations which occur when a single nucleotide is changed in a genetic sequence. These variations can be exploited for use as genetic markers and extrapolated for use in identification assays. The limits of COI-based identification were assessed by defining operational taxonomic units (OTUs), which were established based on phylogenetic clustering of COI amplicons (Frey *et al.*, 2013). Minimal sets of 3 - 4 SNPs were described to differentiate each OTU. The results of this study showed that 10% of tephritid species could not be accurately identified to species level using the diagnostic SNP panels developed. Most of these undifferentiable species were closely related or formed part of species complexes, including the *Bactrocera dorsalis* species complex and the *Ceratitis* FARQ complex.

2.6.2.3. Real-time PCR

The COI barcoding region was assessed for utility in real-time PCR to differentiate between four economically important fruit fly species threatening New Zealand, namely; *C. capitata*, *Bactrocera tyroni* (Froggatt), *B. invadens*, and *Dirioxa pornia* (Walker) (Dhami *et al.*, 2016). Regions of variation in COI between species were selected as targets of probes and primers. TaqMan probe and primer-based assays were developed to identify each of the four species. Although the assays were able to unambiguously differentiate the four species mentioned above, closely related species such as *C. capitata* and *C. caetrata* and species complexes such as the *B. dorsalis* complex could not be differentiated. Thus, the success of this assay is reduced when target species share high identity (Dhami *et al.*, 2016).

2.6.2.4. Microfluidic array

Jiang *et al.* (2016) developed a species-specific probe and primer qPCR assay to identify 27 tephritid pests (including *C. capitata*, *C. cosyra*, *C. rosa* and *B. dorsalis*) based on the COI barcoding region. The assay makes use of a Dynamic Array Integrated Fluidic circuit chip, a real-time PCR system, and 27 primer/probe combinations, with rapid results available in just over seven hours. The assay was successful in identifying 27 fruit flies of interest but had trouble in differentiating members of the *Bactrocera dorsalis* complex. As only one member of the FARQ complex (*C. rosa*) was included in the assay, the applicability for differentiation of this species complex is not yet determined. The only drawback of this proposed diagnostic assay is the cost of the specialised equipment needed that is currently unavailable in all quarantine laboratories as well as the operating costs would be too expensive to run only a few samples at a time (Jiang *et al.*, 2016). The delay caused by waiting for a sample number threshold before running diagnostics could significantly increase the time for species identification if only a few samples are intercepted.

2.6.2.5. LAMP assays

Loop-mediated isothermal amplification or LAMP assays make use of four primers and a strand displacing DNA polymerase, allowing amplification to take place at a constant temperature (Notomi *et al.*, 2015). This type of assay can take place in a single reaction and is faster and more sensitive than a conventional PCR. However, LAMP assays cannot detect multiple species at the same time, as this type of assay cannot be used in a multiplex format. Dermauw *et al.* (2022) designed LAMP primers which targeted COI for *C. capitata*, *C. cosyra* (group 1), and the members of the FARQ complex. The *C. capitata* LAMP primers were specific and did not amplify other tephritid species. Though, it is unlikely that the assay will be able to differentiate between *C. capitata* and the closely related *C. caetrata* due to high sequence similarity in COI. The *C. cosyra* (group 1) primers did not amplify other tephritid species or *C. cosyra* (group 2) specimen. The FARQ primers were specific and amplified only *C. rosa*, *C. quilicii*, *C. anonae* and *C. fasciventris*. Similarly, LAMP primers have also been developed to detect members of the *B. dorsalis* complex (Blaser *et al.*, 2018). It is noted in these studies that the LAMP assays are prone to false-positive and false-negative results.

2.6.3. Broad detection assays

Jiang *et al.* (2018) targeted cytochrome oxidase II (COII) for use in a broad detection assay to detect a wide range of Tephritidae of economic importance to China. A single set of degenerate primers were designed in a region of COII that was conserved between economically important tephritid flies. This target region was amplified in all fruit flies of economic concern and showed no amplification in fruit flies of no concern. The assay was validated to detect 40 fruit fly species in an endpoint PCR, including *C. capitata*, *C. cosyra*, *C. rosa* and *B. dorsalis*. While broad detection assays are highly valuable for detecting the presence of a wider group of economically important species, it does not offer species-level identification. Thus, discrimination between quarantine and nonquarantine pests becomes difficult for countries where regulations differ.

2.6.4. Microsatellite markers

Identification and differentiation between members of the *Ceratitis* FARQ complex was undergone using a set of 16 microsatellite markers which produced five distinct groups, *C. rosa* R1, *C. rosa* R2 (*C. quilicii*), *C. fasciventris* F1, *C. fasciventris* F2, and *C. anonae* (Delatte *et al.*, 2013; Virgilio *et al.*, 2013). The use of a panel of 16 microsatellite markers is time-consuming and expensive. This panel was simplified to include only six microsatellite markers which can accurately differentiate between *C. rosa* and *C. quilicii*; however, when other species are queried against this simplified assay, false-positive results are incurred (Virgilio *et al.*, 2019). While both the 16 and six microsatellite assays are useful in identifying members of the FARQ complex, they are not yet optimised for simultaneous identification or routine use.

2.6.5. Genome-wide SNPs

The *Ceratitis* FARQ complex was investigated in phylogenetic analyses using both mitochondrial data and genome-wide SNPs. Mitochondrial protein coding genes could not resolve species in the *Ceratitis* FARQ

complex as monophyletic clades in a similar fashion to previous studies (Virgilio *et al.*, 2008; De Meyer *et al.*, 2016). However, phylogenetic trees reconstructed using 785 484 genome-wide SNPs recovered from high-throughput sequencing data could reliably resolve the four species into monophyletic groups (Fig. 2.3.) (Zhang *et al.*, 2021). This is the first time the *Ceratitis* FARQ complex has been separated into individual taxa phylogenetically.

2.6.6. Current state of molecular identification in tephritid fruit flies

To date, there is no single molecular tool that can identify the five economically important tephritid species in South Africa to species level simultaneously, timeously, and at a low cost. While multiple molecular tools exist for Tephritidae identification, these tools come with varying limitations. Constant improvements are being made in the resolution of tephritid species as a result of increasing data availability, which goes hand in hand with developing species identification tools which are more accurate and reliable than their predecessors (Virgilio *et al.*, 2019). The existing molecular identification and diagnostic tools evaluated above are dependable frameworks on which further research can build and benefit (Barr *et al.*, 2006), to improve and update the accuracy and reliability of tephritid identification.

2.7. Availability of data

At the onset of this study, in January 2021, the complete mitochondrial genomes of *C. capitata* and *B. dorsalis* were publicly available and accessible from the National Centre for Biotechnology Information (NCBI) GenBank database (Yu *et al.*, 2007; Papanicolaou *et al.*, 2016). Similarly, a segment of COI of each of the five species under study was accessible from NCBI and Barcode of Life (BOLD), as well as reference sequences of NAD6, NAD4, NAD1, and 12S and 16S rRNA for all species except for the recently described *C. quilicii*. The complete mitochondrial genomes of the other three fruit flies of interest in this study, *C. cosyra*, *C. rosa* and *C. quilicii,* were published later during the course of 2021 while the current study was already underway (Drosopoulou *et al.*, 2021; Zhang *et al.*, 2021). The complete genome of *C. capitata* (Papanicolaou *et al.*, 2016) and *B. dorsalis* was sequenced and assembled in 2016 and 2019, respectively. The shotgun sequence data for each species is accessible through GenBank as unplaced genomic scaffolds with putative annotations provided by the NCBI's Gnomon gene prediction tool.

2.7.1. High Throughput Sequencing – a way forward?

High throughput sequencing (HTS) offers the unique advantage of in-depth analysis of mitochondrial and genomic data. For insect identification, current molecular diagnostic and identification techniques are mostly based on the mitochondrial genome and well conserved protein-coding regions within it. However, it has become evident that in the case of closely related species and species complexes, these regions do not offer enough variation to distinguish between species (Zhang *et al.*, 2021). Thus, the next port of call would be to look at less conserved mitochondrial regions or genomic regions for differentiation of these closely related species. HTS can be used to explore a range of genomic and mitogenomic differences between species as it

has high sensitivity and can detect low-frequency variants. Accurate species identification relies on the availability of reference libraries and databases from which suitable detection tools can be developed (Littlefair and Clare, 2016). HTS and whole genome sequencing (WGS) are useful for identifying potential molecular markers in non-model insects (Perry and Rowe, 2011; Wachi, Matsubayashi and Maeto, 2018). Most importantly, species-specific HTS and WGS data generated can be used as a basis for further research. For instance, data mining of existing genome data of the tephritid fly *Lucilia cuprina* (Wiedemann) has resulted in an embryonic sexing system for improving the SIT program for this fly (Yan and Scott, 2015).

2.8. The road ahead in Tephritidae identification

2.8.1. Variable mitochondrial regions

Relatively conserved regions in the mitochondria such as protein coding genes (PCG's) including, COI, COII, NAD1-6, ribosomal RNA's (rRNA's), and transfer RNA's (tRNA's) have been the focus of attention in terms of identification assays and phylogenetic analyses. While these studies have improved the resolution of species and species complexes boundlessly, they have not been able to address the gap identified in this study – accurately differentiating between *C. capitata*, *B. dorsalis*, *C. rosa*, *C. quilicii* and *C. cosyra* using a single molecular assay. Variable mitochondrial intergenic spacers (IGS) are an emerging resource which could potentially contribute to the identification of closely related species (Smith and Bush, 1997; da Silva *et al.*, 2009). These intergenic regions are noncoding and typically flank coding domains such as tRNA's. Intergenic regions can be phylogenetically informative based on simply the presence or absence of an IGS as well as length variations between species (Xu, Zhou and Wan, 2022). The difference in length of IGS can vary considerably between closely related species (McClelland, Petersen and Welsh, 1992). A technique known as "tRNA intergenic spacer PCR", also called tDNA-PCR, has previously demonstrated competence in differentiating between species of various bacteria (Baele *et al.*, 2000). This technique exploits conserved, flanking tRNA's for primer design to amplify the variable IGS. PCR products can then be separated using capillary electrophoresis for identification. The largest intergenic spacer in *C. quilicii* and *C. rosa* is located between $tRNA^{ile}$ and $tRNA^{gln}$ (Drosopoulou *et al.*, 2021). It is proposed that intergenic spacers could play a role in insect evolution and are potentially indicative of differing environmental pressures leading to evolutionary events and speciation. The mitochondrial intergenic spacer between $tRNA^{ile}$ and $tRNA^{gln}$ has been emphasised as a potential marker for use in further phylogenomic and species identification efforts due to the size variation of this region between species (da Silva *et al.*, 2009; Xu, Zhou and Wan, 2022).

2.8.2. Genomic regions of interest

Research into genomic DNA is gaining momentum for species differentiation tools. Genome-wide SNPs have had the greatest success in distinguishing between closely related, cryptic species *C. rosa* and *C. quilicii* (Zhang *et al.*, 2021). It has been proposed that the shifting of host plant use (for feeding or oviposition) is a mechanism for the formation of reproductive isolation by adaptation to new ecological niches and a driving force of speciation. Speciation due to host shifting in Tephritidae has already been demonstrated within the

genus *Rhagoletis* with *R. pomonella* (Walsh) (Forbes *et al.*, 2009). The mechanisms involved in host plant selection and recognition include both chemosensation, such as gustatory and olfactory receptors, as well as visual stimuli (Feuda *et al.*, 2016; Kasubuchi *et al.*, 2018). Adaptation to new environments requires taste-related systems to adapt in order to identify potential chemical threats in the new environment (Dweck and Carlson, 2020), which are a common defence response from plants under stress. Oviposition and food-seeking behaviours are also driven by olfactory senses. Hence, odour and gustatory receptors have been associated with the adaptation to new ecological niches in *Drosophila* (Hickner *et al.*, 2016; Shaw *et al.*, 2019). Host fruit detection for oviposition of the medfly (*C. capitata*) is partially led by visual capabilities, in which opsins play a major role (Papanicolaou *et al.*, 2016). Opsin evolution in Diptera is highly variable, demonstrating evidence for positive selection of photic niche and adaptive evolution of UV-sensitive opsins (Rh3/Rh4) in day-flying insects (Feuda *et al.*, 2016, 2021). These adaptations are associated with host plant detection and wing colour/pattern recognition conferring the ability to respond to different wavelengths of light. Evidence for the functional utility of opsins expands beyond the visual system and into chemosensation, with opsins Rh1, Rh4 and Rh7 involved in gustatory preferences (Leung *et al.*, 2020).

2.8.3. Tools for thought

While a great variety of molecular identification tools exist, few have been utilised for species differentiation of tephritid flies with varying success. One promising technique is high resolution melt analysis (HRM), whereby the melting behaviour of PCR amplicons is explored, and differences in DNA sequence result in differences in melting temperature (Tm). Morphologically cryptic mealybug species, which are notoriously misidentified, have been successfully differentiated using HRM on a small (158 bp) fragment of COI (Wetten, Campbell and Allainguillaume, 2016). Similarly, morphologically identical *Anopheles* species were accurately distinguished with HRM based on SNPs present in an rDNA intergenic spacer (Zianni *et al.*, 2013). HRM has also proven to be effective in determining which species are present in a mixed sample and can be used in conjunction with multiple primers amplifying different species-specific products in a multiplex format (Elkins, Perez and Sweetin, 2016). The use of end-point multiplex PCR for diagnostic purposes is more commonly used in cases where there are potential quarantine implications as the equipment needed is widely available in most laboratories. This type of assay has been developed for the differentiation of *D. suzukii* (Matsumura) from other *Drosophila* species at PoE for time-sensitive quarantine decisions (Murphy *et al.*, 2015). Multiplex PCR has shown promising results for the identification of Tephritidae, successfully differentiating between wild and sterile *Bactrocera tyroni* populations with nine microsatellite markers in a single reaction (Chen, Dominiak and O'Rourke, 2016) and distinguishing *R. ceraci* Linnaeus from other *Rhagoletis* species (Barr *et al.*, 2021). Overall, both HRM and multiplex PCR demonstrate proficiency in differentiating between closely related and morphologically cryptic species allowing accurate identification that is not life-stage specific.

2.9. Conclusion

The state of current literature pertaining to fruit fly identification is an ever-growing and pertinent topic worldwide, particularly for major export and import countries of fresh produce. The value of resources which aid in the detection and identification of fruit fly pests of quarantine significance cannot be understated. In the South African context, the availability of molecular resources would considerably improve fruit fly monitoring and management practices, in addition to supporting decisions made in packhouses and at PoE. Such molecular tools can significantly reduce the turnaround time for quarantine decisions, which in turn reduces costs and increases the shelf life and quality of fresh fruit consignments. Prior to this study, there was no single molecular assay that could accurately differentiate between the five fruit flies of economic importance to the South African fresh fruit export industry: *C. capitata*, *C. cosyra*, *C. rosa*, *C. quilicii* and *B. dorsalis*. The present study aimed to develop such an assay using HTS data, and genomic and mitogenomic variation between these five fruit fly species were exploited for identification assays using multiple molecular methods (HRM analysis and Multiplex PCR). The current study will not only benefit the South African fresh fruit industry through the identification of phytosanitary pests but provide valuable data and insight into the problem of tephritid fruit flies.

2.10. References

Armstrong, K. F. and Ball, S. L. (2005) 'DNA barcodes for biosecurity: invasive species identification', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1462), pp. 1813–1823. doi: 10.1098/rstb.2005.1713.

Armstrong, K. F., Cameron, C. M. and Frampton, E. R. (1997) 'Fruit fly (Diptera: Tephritidae) species identification: a rapid molecular diagnostic technique for quarantine application', *Bulletin of Entomological Research*, 87(2), pp. 111–118. doi: 10.1017/S0007485300027243.

Badii, K. B. *et al.* (2015) 'Review of the pest status, economic impact and management of fruit-infesting flies (Diptera: Tephritidae) in Africa', *African Journal of Agricultural Research*, 10(12), pp. 1488–1498. doi: 10.5897/AJAR2014.9278.

Baele, M. *et al.* (2000) 'Application of tRNA intergenic spacer PCR for identification of *Enterococcus* species', *Journal of Clinical Microbiology*, 38(11), pp. 4201–4207.

Balmès, V. and Mouttet, R. (2017) 'Development and validation of a simplified morphological identification key for larvae of tephritid species most commonly intercepted at import in Europe', *EPPO Bulletin*, 47(1), pp. 91–99. doi: 10.1111/epp.12369.

Barnes, B. and Venter, J. (2006) 'The South African fruit fly action plan: area-wide suppression and exotic species surveillance', in *Proceedings of the 7th International Symposium on Fruit Flies of Economic Importance*. Salvador, Brazil, pp. 271–283.

Barr, N. *et al.* (2006) 'Molecular diagnostics of economically important *Ceratitis* fruit fly species (Diptera: Tephritidae) in Africa using PCR and RFLP analyses', *Bulletin of Entomological Research*, 96, pp. 505–521. doi: 10.1079/BER2006452.

Barr, N. B. *et al.* (2012) 'Molecular Identification of *Ceratitis capitata* (Diptera: Tephritidae) Using DNA Sequences of the COI Barcode Region', *Annals of the Entomological Society of America*, 105(2), pp. 339–350. doi: 10.1603/AN11100.

Barr, N. B. *et al.* (2021) 'Using the rDNA Internal Transcribed Spacer 1 to Identify the Invasive Pest *Rhagoletis cerasi* (Diptera: Tephritidae) in North America', *Journal of Economic Entomology*. Edited by K. Shelby, 114(1), pp. 360–370. doi: 10.1093/jee/toaa281.

Bezzi, M. (1920) 'Notes on the Ethiopian fruit flies of the family Trypaneidae, other than Dacus', *Bulletin of Entomological Research*, 10, pp. 211–272.

Blacket, M. J., Semeraro, L. and Malipatil, M. B. (2012) 'Barcoding Queensland Fruit Flies (*Bactrocera tryoni*): impediments and improvements', *Molecular Ecology Resources*, 12(3), pp. 428–436. doi: 10.1111/j.1755-0998.2012.03124.x.

Blaser, S. *et al.* (2018) 'From laboratory to point of entry: development and implementation of a loop-mediated isothermal amplification (LAMP)-based genetic identification system to prevent introduction of quarantine insect species', *Pest Management Science*, 74(6), pp. 1504–1512. doi: 10.1002/ps.4866.

Boller, E. F. *et al.* (1976) 'Incompatible races of European cherry fruit fly, *Rhagoletis cerasi* (Diptera: Tephritidae), their origin and potential use in biological control', *Entomologia Experimentalis et Applicata*, 20(3), pp. 237–247. doi: 10.1111/j.1570-7458.1976.tb02640.x.

Boykin, L. M. *et al.* (2012) 'Species Delimitation and Global Biosecurity', *Evolutionary Bioinformatics*, 8, p. EBO.S8532. doi: 10.4137/EBO.S8532.

Bruzzese, D. J. *et al.* (2022) 'Testing the potential contribution of *Wolbachia* to speciation when cytoplasmic incompatibility becomes associated with host-related reproductive isolation', *Molecular Ecology*, 31(10), pp. 2935–2950. doi: 10.1111/mec.16157.

Chen, Y., Dominiak, B. C. and O'Rourke, B. A. (2016) 'A single multiplex PCR reaction for distinguishing strains of Queensland fruit fly *Bactrocera tryoni* (Diptera: Tephritidae)', *Austral Entomology*, 55(3), pp. 316–323. doi: 10.1111/aen.12190.

Christenson, L. D. and Foote, R. H. (1960) 'Biology of Fruit Flies', *Annual Review of Entomology*, 5(1), pp. 171–192. doi: 10.1146/annurev.en.05.010160.001131.

Copeland, R. S. *et al.* (2006) 'Geographic Distribution, Host Fruit, and Parasitoids of African Fruit Fly Pests *Ceratitis anonae*, *Ceratitis cosyra*, *Ceratitis fasciventris*, and *Ceratitis rosa* (Diptera: Tephritidae) in Kenya', *Annals of the Entomological Society of America*, 99(2), pp. 261–278.

Cugala, D. *et al.* (2014) 'Assessment of ripening stages of Cavendish dwarf bananas as host or non-host to *Bactrocera invadens*', *Journal of Applied Entomology*, 138(6), pp. 449–457. doi: 10.1111/jen.12045.

da Silva, N. M. *et al.* (2009) 'Characterization of mitochondrial control region, two intergenic spacers and tRNAs of *Zaprionus indianus* (Diptera: Drosophilidae)', *Genetica*, 137(3), pp. 325–332. doi: 10.1007/s10709-009-9396-5.

Delatte, H. *et al.* (2013) 'Isolation and characterisation of sixteen microsatellite markers cross-amplifying in a complex of three African agricultural pests (*Ceratitis rosa*, *C. anonae* and *C. fasciventris*, Diptera: Tephritidae)', *Conservation Genetics Resources*, 5(1), pp. 31–34. doi: 10.1007/s12686-012-9722-6.

De Meyer, M. (2001) 'On the identity of the Natal fruit fly *Ceratitis rosa* Karsch (Diptera, Tephritidae)', *Bulletin de l'Institut Royal des Sciences Naturelles de Belgique, Entomologie*, 71, pp. 55–62.

De Meyer, M. *et al.* (2002) *Annotated check list of host plants for Afrotropical fruit flies (Diptera: Tephritidae) of the genus Ceratitis*, *MRAC*.

De Meyer, M. *et al.* (2016) 'Description of new *Ceratitis* species (Diptera: Tephritidae) from Africa, or how morphological and DNA data are complementary in discovering unknown species and matching sexes', *European Journal of Taxonomy*, (233). doi: 10.5852/ejt.2016.233.

De Meyer, M. and Freidberg, A. (2005) 'Revision of the Subenus *Ceratitis (pterandrus)* Bezzi (Diptera: Tephritidae)', *Biotaxonomy of Tephritoidea*, 35(6), pp. 197–315.

Dermauw, W. *et al.* (2022) 'A loop-mediated isothermal amplification (LAMP) assay for rapid identification of *Ceratitis capitata* and related species', *Current Research in Insect Science*, 2, p. 100029. doi: 10.1016/j.cris.2022.100029.

De Villiers, M. *et al.* (2013) 'The Distribution, Relative Abundance, and Seasonal Phenology of *Ceratitis capitata* , *Ceratitis rosa* , and *Ceratitis cosyra* (Diptera: Tephritidae) in South Africa', *Environmental Entomology*, 42(5), pp. 831–840. doi: 10.1603/EN12289.

Dhami, M. K. *et al.* (2016) 'A real-time PCR toolbox for accurate identification of invasive fruit fly species', *Journal of Applied Entomology*, 140(7), pp. 536–552. doi: 10.1111/jen.12286.

Drew, R. A. I. and Hancock, D. L. (1994) 'The *Bactrocera dorsalis* complex of fruit flies (Diptera: Tephritidae: Dacinae) in Asia', *Bulletin of entomological research supplement series*, 2, pp. 1–68. doi: 10.1017/S1367426900000278.

Drew, R. and Romig, M. (2013) *Tropical fruit flies (Tephritidae Dacinae) of South-East Asia: Indomalaya to North-West Australasia*, *CABI*.

Drosopoulou, E. *et al.* (2021) 'The complete mitochondrial genomes of *Ceratitis rosa* and *Ceratitis quilicii* , members of the *Ceratitis* FAR species complex (Diptera: Tephritidae)', *Mitochondrial DNA Part B*, 6(3), pp. 1039–1041. doi: 10.1080/23802359.2021.1899073.

Duyck, P. F., David, P. and Quilici, S. (2004) 'A review of relationships between interspecific competition and invasions in fruit flies (Diptera: Tephritidae)', *Ecological Entomology*, 29(5), pp. 511–520. doi: 10.1111/j.0307-6946.2004.00638.x.

Dweck, H. K. M. and Carlson, J. R. (2020) 'Molecular Logic and Evolution of Bitter Taste in *Drosophila*', *Current Biology*, 30(1), pp. 17-30.e3. doi: 10.1016/j.cub.2019.11.005.

Ekesi, S. and Billah, M. K. (eds) (2006) *Field guide to the management of economically important tephritid fruit flies in Africa*. 2nd edn. Nairobi, Kenya: ICIPE Science Press.

Elkins, K. M., Perez, A. C. U. and Sweetin, K. C. (2016) 'Rapid and inexpensive species differentiation using a multiplex real-time polymerase chain reaction high-resolution melt assay', *Analytical Biochemistry*, 500, pp. 15–17. doi: 10.1016/j.ab.2016.01.013.

Erbout, N., De Meyer, M. and Lens, L. (2008) 'Hybridization between two polyphagous fruit-fly species (Diptera: Tephritidae) causes sex-biased reduction in developmental stability', *Biological Journal of the Linnean Society*, (93), pp. 579–588.

Feuda, R. *et al.* (2016) 'Conservation, Duplication, and Divergence of Five Opsin Genes in Insect Evolution', *Genome Biology and Evolution*, 8(3), pp. 579–587. doi: 10.1093/gbe/evw015.

Feuda, R. *et al.* (2021) 'Phylogenomics of Opsin Genes in Diptera Reveals Lineage-Specific Events and Contrasting Evolutionary Dynamics in *Anopheles* and *Drosophila*', *Genome Biology and Evolution*. Edited by A. Betancourt, 13(8). doi: 10.1093/gbe/evab170.

Follett, P. A. and Snook, K. (2013) 'Cold Storage Enhances the Efficacy and Margin of Security in Postharvest Irradiation Treatments Against Fruit Flies (Diptera: Tephritidae)', *Journal of Economic Entomology*, 106(5), pp. 2035–2042. doi: 10.1603/EC13197.

Forbes, A. A. *et al.* (2009) 'Sequential Sympatric Speciation Across Trophic Levels', *Science*, 323(5915), pp. 776–779. doi: 10.1126/science.1166981.

Frey, J. E. *et al.* (2013) 'Developing diagnostic SNP panels for the identification of true fruit flies (Diptera: Tephritidae) within the limits of COI-based species delimitation', *BMC Evolutionary Biology*, 13(1), p. 106. doi: 10.1186/1471-2148-13-106.

Funk, D. . and Omland, K. . (2003) 'Species-Level Paraphyly and Polyphyly: Frequency, Causes, and Consequences, with Insights from Animal Mitochondrial DNA', *Annual Review of Ecology, Evolution and Systematics*, 34, pp. 397–423.

Garzón-Orduña, I. J. *et al.* (2020) 'Implementing Low-Cost, High Accuracy DNA Barcoding From Single Molecule Sequencing to Screen Larval Tephritid Fruit Flies Intercepted at Ports of Entry', *Annals of the Entomological Society of America*. Edited by K. Gaddis, 113(4), pp. 288–297. doi: 10.1093/aesa/saz071.

Grout, T. G. *et al.* (2011) 'Cold Treatment of *Ceratitis capitata* (Diptera: Tephritidae) in Oranges Using a Larval Endpoint', *Journal of Economic Entomology*, 104(4), pp. 1174–1179. doi: 10.1603/EC10434.

Grout, T. G. (2015) 'The Status of Citrus IPM in South Africa', *Acta Horticulturae*. Edited by B. Sabater-Muñoz, (1065), pp. 1091–1095. doi: 10.17660/ActaHortic.2015.1065.137.

Hebert, P. D. N. *et al.* (2003) 'Biological identifications through DNA barcodes', *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1512), pp. 313–321. doi: 10.1098/rspb.2002.2218.

Hickner, P. V. *et al.* (2016) 'The making of a pest: Insights from the evolution of chemosensory receptor families in a pestiferous and invasive fly, *Drosophila suzukii*', *BMC Genomics*, 17(1), p. 648. doi: 10.1186/s12864-016-2983-9.

ISPM-35 (2018) 'Systems approach for pest risk management of fruit flies (Tephritidae)', *IPPC, FAO*. Available at: www.ippc.int.

Jiang, F. *et al.* (2014) 'Existence of species complex largely reduced barcoding success for invasive species of Tephritidae: a case study in *Bactrocera* spp.', *Molecular Ecology Resources*, 14(6), pp. 1114–1128. doi: 10.1111/1755-0998.12259.

Jiang, F. *et al.* (2016) 'A high-throughput detection method for invasive fruit fly (Diptera: Tephritidae) species based on microfluidic dynamic array', *Molecular Ecology Resources*, 16(6), pp. 1378–1388. doi: 10.1111/1755-0998.12542.

Jiang, F. *et al.* (2018) 'A conserved motif within cox 2 allows broad detection of economically important fruit flies (Diptera: Tephritidae)', *Scientific Reports*, 8(1), p. 2077. doi: 10.1038/s41598-018-20555-2.

Kasubuchi, M. *et al.* (2018) 'Insect taste receptors relevant to host identification by recognition of secondary metabolite patterns of non-host plants', *Biochemical and Biophysical Research Communications*, 499(4), pp. 901–906. doi: 10.1016/j.bbrc.2018.04.014.

Keeling, M. J., Jiggins, F. M. and Read, J. M. (2003) 'The invasion and coexistence of competing *Wolbachia* strains', *Heredity*, pp. 382–388.

Leblanc, L. *et al.* (2015) 'A phylogenetic assessment of the polyphyletic nature and intraspecific color polymorphism in the *Bactrocera dorsalis* complex (Diptera, Tephritidae)', *ZooKeys*, 540, pp. 339–367. doi: 10.3897/zookeys.540.9786.

Leung, N. Y. *et al.* (2020) 'Functions of Opsins in *Drosophila* Taste', *Current Biology*, 30(8), pp. 1367-1379.e6. doi: 10.1016/j.cub.2020.01.068.

Li, B. *et al.* (2009) 'Potential Geographical Distributions of the Fruit Flies *Ceratitis capitata*, *Ceratitis cosyra*, and *Ceratitis rosa* in China', *Journal of Economic Entomology*, 102(5), pp. 1781–1790.

Littlefair, J. E. and Clare, E. L. (2016) 'Barcoding the food chain: from Sanger to high-throughput sequencing', *Genome*. Edited by A. Naaum, 59(11), pp. 946–958. doi: 10.1139/gen-2016-0028.

Lukhtanov, V. A. (2019) 'Species Delimitation and Analysis of Cryptic Species Diversity in the XXI Century', *Entomological Review*, 99(4), pp. 463–472. doi: 10.1134/S0013873819040055.

Manrakhan, A. (2020) 'Fruit Fly', in Grout, T. (ed.) *Integrated Production Guidelines for export citrus.* Nelspruit, South Africa: Citrus Research International, pp. 1–10.

Manrakhan, A., Venter, J. H. and Hattingh, V. (2015) 'The progressive invasion of *Bactrocera dorsalis* (Diptera: Tephritidae) in South Africa', *Biological Invasions*, 17(10), pp. 2803–2809. doi: 10.1007/s10530-015-0923-2.

Mateos, M. *et al.* (2020) '*Wolbachia pipientis* Associated With Tephritid Fruit Fly Pests: From Basic Research to Applications', *Frontiers in Microbiology*, 11. doi: 10.3389/fmicb.2020.01080.

McClelland, M., Petersen, C. and Welsh, J. (1992) 'Length polymorphisms in tRNA intergenic spacers detected by using the polymerase chain reaction can distinguish Streptococcal strains and species', *Journal of Clinical Microbiology*, 30(6), pp. 1499–1504.

Morrow, J. L. *et al.* (2014) 'Tropical tephritid fruit fly community with high incidence of shared *Wolbachia strains* as platform for horizontal transmission of endosymbionts', *Environmental Microbiology*, 16(12), pp. 3622–3637. doi: 10.1111/1462-2920.12382.

Murphy, K. A. *et al.* (2015) 'Using comparative genomics to develop a molecular diagnostic for the identification of an emerging pest *Drosophila suzukii*', *Bulletin of Entomological Research*, 105(3), pp. 364–372. doi: 10.1017/S0007485315000218.

Mwatawala, M. *et al.* (2015) 'Niche partitioning among two *Ceratitis rosa* morphotypes and other *Ceratitis* pest species (Diptera, Tephritidae) along an altitudinal transect in Central Tanzania', *ZooKeys*, 540, pp. 429–442. doi: 10.3897/zookeys.540.6016.

Nardi, F. *et al.* (2005) 'Population structure and colonization history of the olive fly, *Bactrocera oleae* (Diptera, Tephritidae)', *Molecular Ecology*, 14(9), pp. 2729–2738. doi: 10.1111/j.1365-294X.2005.02610.x.

Notomi, T. *et al.* (2015) 'Loop-mediated isothermal amplification (LAMP): principle, features, and future prospects', *Journal of Microbiology*, 53(1), pp. 1–5. doi: 10.1007/s12275-015-4656-9.

Papanicolaou, A. *et al.* (2016) 'The whole genome sequence of the Mediterranean fruit fly, *Ceratitis capitata* (Wiedemann), reveals insights into the biology and adaptive evolution of a highly invasive pest species', *Genome Biology*, 17(1), p. 192. doi: 10.1186/s13059-016-1049-2.

Perry, J. C. and Rowe, L. (2011) 'Rapid Microsatellite Development for Water Striders by Next-Generation Sequencing', *Journal of Heredity*, 102(1), pp. 125–129. doi: 10.1093/jhered/esq099.

Prezotto, L. F. *et al.* (2017) '*Wolbachia* strains in cryptic species of the *Anastrepha fraterculus* complex (Diptera, Tephritidae) along the Neotropical Region', *Systematic and Applied Microbiology*, 40(1), pp. 59–67. doi: 10.1016/j.syapm.2016.11.002.

Riegler, M. and Stauffer, C. (2008) '*Wolbachia* infections and superinfections in cytoplasmically incompatible populations of the European cherry fruit fly *Rhagoletis cerasi* (Diptera, Tephritidae)', *Molecular Ecology*, 11(11), pp. 2425–2434. doi: 10.1046/j.1365-294X.2002.01614.x.

Schutze, M. K. *et al.* (2015) 'Synonymization of key pest species within the *Bactrocera dorsalis* species complex (Diptera: Tephritidae): taxonomic changes based on a review of 20 years of integrative morphological, molecular, cytogenetic, behavioural and chemoecological data', *Systematic Entomology*, 40(2), pp. 456–471. doi: 10.1111/syen.12113.

Schutze, M. K. *et al.* (2017) 'Tephritid Integrative Taxonomy: Where We Are Now, with a Focus on the Resolution of Three Tropical Fruit Fly Species Complexes', *Annual Review of Entomology*, 62(1), pp. 147–164. doi: 10.1146/annurev-ento-031616-035518.

Shaw, K. H. *et al.* (2019) 'Molecular and Functional Evolution at the Odorant Receptor Or22 Locus in *Drosophila melanogaster*', *Molecular Biology and Evolution*. Edited by P. Wittkopp, 36(5), pp. 919–929. doi: 10.1093/molbev/msz018.

Smith, J. . and Bush, G. . (1997) 'Phylogeny of the Genus *Rhagoletis* (Diptera: Tephritidae) Inferred from DNA Sequences of Mitochondrial Cytochrome Oxidase II', *Molecular Phylogenetics and Evolution*, 7(1), pp. 33–43.

Tan, K. H. *et al.* (2013) 'Comparison of methyl eugenol metabolites, mitochondrial COI, and rDNA sequences of *Bactrocera philippinensis* (Diptera: Tephritidae) with those of three other major pest species within the *dorsalis* complex', *Applied Entomology and Zoology*, 48(3), pp. 275–282. doi: 10.1007/s13355-013-0183-5.

Tanga, C. M. *et al.* (2018) 'Risk assessment and spread of the potentially invasive *Ceratitis rosa* Karsch and *Ceratitis quilicii* De Meyer, Mwatawala & Virgilio sp. Nov. using life-cycle simulation models: Implications for phytosanitary measures and management', *PLOS ONE*. Edited by N. T. Papadopoulos, 13(1), p. e0189138. doi: 10.1371/journal.pone.0189138.

Virgilio, M. *et al.* (2008) 'Molecular evaluation of nominal species in the *Ceratitis fasciventris*, *C. anonae*, *C. rosa* complex (Diptera: Tephritidae)', *Molecular Phylogenetics and Evolution*, 48(1), pp. 270–280. doi: 10.1016/j.ympev.2008.04.018.

Virgilio, M. *et al.* (2012) 'Identifying Insects with Incomplete DNA Barcode Libraries, African Fruit Flies (Diptera: Tephritidae) as a Test Case', *PLoS ONE*. Edited by D. Steinke, 7(2), p. e31581. doi: 10.1371/journal.pone.0031581.

Virgilio, M. *et al.* (2013) 'Cryptic diversity and gene flow among three African agricultural pests: *Ceratitis rosa* , *Ceratitis fasciventris* and *Ceratitis anonae* (Diptera, Tephritidae)', *Molecular Ecology*, 22(9), pp. 2526–2539. doi: 10.1111/mec.12278.

Virgilio, M. *et al.* (2017) 'The complex case of *Ceratitis cosyra* (Diptera: Tephritidae) and relatives. A DNA barcoding perspective', *Journal of Applied Entomology*, 141(10), pp. 788–797. doi: 10.1111/jen.12379.

Virgilio, M. *et al.* (2019) 'An integrated diagnostic setup for the morphological and molecular identification of the *Ceratitis* FAR complex (*C. anonae, C. fasciventris, C. rosa, C. quilicii*, Diptera, Tephritidae)', *Bulletin of Entomological Research*, 109(3), pp. 376–382. doi: 10.1017/S0007485318000615.

Virgilio, M., White, I. and De Meyer, M. (2014) 'A set of multi-entry identification keys to African frugivorous flies (Diptera, Tephritidae)', *ZooKeys*, 428, pp. 97–108. doi: 10.3897/zookeys.428.7366.

Wachi, N., Matsubayashi, K. W. and Maeto, K. (2018) 'Application of next-generation sequencing to the study of non-model insects', *Entomological Science*, 21(1), pp. 3–11. doi: 10.1111/ens.12281.

Waugh, J. (2007) 'DNA barcoding in animal species: progress, potential and pitfalls', *BioEssays*, 29(2), pp. 188–197. doi: 10.1002/bies.20529.

Weeks, A. R., Reynolds, T. K. and Hoffmann, A. A. (2002) '*Wolbachia* dynamics and host effects: what has (and has not) been demonstrated?', *Trends in Ecology & Evolution*, 17(6), pp. 257–262. doi: 10.1016/S0169-5347(02)02480-1.

Wetten, A., Campbell, C. and Allainguillaume, J. (2016) 'High-resolution melt and morphological analyses of mealybugs (Hemiptera: Pseudococcidae) from cacao: tools for the control of, *Cacao swollen shoot virus* spread', *Pest Management Science*, 72(3), pp. 527–533. doi: 10.1002/ps.4017.

Whatson, M. (2020) 'Decision To Revise Import Requirements for the Importation of Fresh Citrus From South Africa Into the United States', *Federal Registar*, 85(215).

White, I. M. and Elson-Harris, M. M. (1992) *Fruit flies of economic significance: their identification and bionomics*. CAB international.

Xu, M., Zhou, S. and Wan, X. (2022) 'Phylogenetic Implication of Large Intergenic Spacers: Insights from a Mitogenomic Comparison of *Prosopocoilus* Stag Beetles (Coleoptera: Lucanidae)', *Animals*, 12(13), p. 1595. doi: 10.3390/ani12131595.

Yan, Y. and Scott, M. J. (2015) 'A transgenic embryonic sexing system for the Australian sheep blow fly *Lucilia cuprina*', *Scientific Reports*, 5(1), p. 16090. doi: 10.1038/srep16090.

Yu, D. J. *et al.* (2007) 'The complete nucleotide sequence of the mitochondrial genome of the oriental fruit fly, *Bactrocera dorsalis* (Diptera: Tephritidae)', *Gene*, 396(1), pp. 66–74. doi: 10.1016/j.gene.2007.02.023.

Zhang, Y. *et al.* (2021) 'Phylogenomic resolution of the *Ceratitis* FARQ complex (Diptera: Tephritidae)', *Molecular Phylogenetics and Evolution*, 161, p. 107160. doi: 10.1016/j.ympev.2021.107160.

Zianni, M. R. *et al.* (2013) 'Rapid Discrimination between, *Anopheles gambiae s.s.* and *Anopheles arabiensis* by High-Resolution Melt (HRM) Analysis', *Journal of Biomolecular Techniques : JBT*, p. jbt.13-2401-001. doi: 10.7171/jbt.13-2401-001.

CHAPTER 3. UTILISATION OF MITOCHONDRIAL INTERGENIC REGION FOR SPECIES DIFFERENTIATION OF FRUIT FLIES (DIPTERA: TEPHRITIDAE) IN SOUTH AFRICA

3.1. Background

Five fruit fly species in the family Tephritidae (Order Diptera) affect fresh fruit production and export in South Africa (Prinsloo and Uys, 2015). Four of these flies are of Afrotropical origin belonging to the genus *Ceratitis* MacLeay; *Ceratitis capitata* (Wiedemann), Mediterranean fruit fly; *Ceratitis cosyra* (Walker), marula fly; *Ceratitis rosa* Karsch, Natal fly; and *Ceratitis quilicii* De Meyer, Mwatawala & Virgilio, the Cape fly (De Meyer, 2001). The latter fruit fly is a newly described species; its current host range and geographic distribution are still being determined (De Meyer *et al.*, 2016). The fifth fruit fly species: *Bactrocera dorsalis* (Hendel), is of Asian origin and invaded the northern areas of South Africa in 2013 (Manrakhan, Venter and Hattingh, 2015). Fruit fly pests cause physical damage to fruit produced in South Africa through oviposition, leaving puncture marks on the skin giving rise to flesh-decay via secondary bacterial or fungal infections, rendering it unmarketable (Steck, 2004; Ekesi and Billah, 2006). Other economic damage is incurred from export market restrictions due to the quarantine status and invasion potential of these flies (Malacrida *et al.*, 2007; Grout, 2015; Mommsen and Bester, 2018; Qin *et al.*, 2018; Louzeiro *et al.*, 2021). While the five fruit fly species focused on in this study are not the only tephritid pests present in South Africa, they are currently the only economically significant fruit fly pests of commercial fresh fruit exported from South Africa. Country-specific phytosanitary certification measures are in place to ensure that consignments containing plant products such as fresh fruit are free from quarantine pests upon arrival at Ports of Entry (PoE) (Whatson, 2020; IPPC, 2021). Export market requirements change constantly based on the absence or presence of pests in both the exporting and importing countries. The European Union (EU) is a significant export market for South Africa and is responsible for up to 46% of fresh fruit exports annually (Fruit South Africa, *2020 Key fruit statistics*). This market has zero-tolerance enforcement for the presence of non-EU Tephritidae, which includes all fruit fly pests in South Africa, except *C. capitata*, which is an established pest in the EU (Bragard *et al.*, 2020).

Therefore, it is necessary to accurately and reliably identify these five fruit fly species through surveillance practices prior to export. The primary goal of fruit fly management is to produce commercial fruit that are free of fruit flies. To facilitate this process, fruit fly management practices are applied before harvest, including pest monitoring, orchard sanitation, and the application of control products. A number of measures are applied after harvest, such as sorting, inspection, and, where necessary disinfestation treatments. Fruit fly surveillance programmes are also in place to detect the presence of exotic species such as *B. dorsalis* in pest-free areas in South Africa (Venter, Baard and Barnes, 2021). The success of these management practices is evaluated through routine monitoring and inspection programmes that provide estimates of population size and are useful in declaring pest-free zones. Such surveillance programmes should accurately identify all fruit flies of economic significance to South Africa to species level (Manrakhan, 2020). Morphological keys for identifying adult fruit flies and third-instar larval specimens have been well developed (Virgilio, White and

De Meyer, 2014; Balmès and Mouttet, 2017). However, morphological identification becomes problematic when specimens are damaged, cryptic species are found, or early life stages are intercepted (Boykin *et al.*, 2012; De Meyer *et al.*, 2016; Balmès and Mouttet, 2017; Virgilio *et al.*, 2019). When cases arise where species cannot be reliably distinguished through morphological methods, the use of molecular diagnostics would be more efficient.

DNA barcoding using cytochrome oxidase I (COI) has been used as a standard DNA marker for species identification. This molecular marker is relatively conserved within the same species, with variation present between different species allowing for identification (Waugh, 2007). This technique has been demonstrated to resolve most species; however, DNA barcoding becomes problematic when closely related and cryptic species are present where interspecific variation is reduced. The COI region of the five fruit flies in this study have previously been investigated for application in species identification. These results were unable to resolve species complexes such as the *Ceratitis* FARQ complex (*Ceratitis fasciventris* (Bezzi), *Ceratitis anonae* Graham, *Ceratitis rosa*, and *Ceratitis quilicii*) and the *Bactrocera dorsalis* complex (Virgilio *et al.*, 2008, 2012; Frey *et al.*, 2013; Jiang *et al.*, 2014; De Meyer *et al.*, 2016; Zhang *et al.*, 2021). A molecular assay that can differentiate between all five species simultaneously would be valuable for routine monitoring and pest surveillance. A multiplex assay has recently been developed for use at PoE in cases of larval interception (Andrews *et al.*, 2022). While this assay is useful for time-sensitive identification matters at PoE, routine pest monitoring and surveillance may benefit from a sequencing-based assay as the resources are readily available at these facilities. A sequencing-based assay will assist in generating a valuable database where any diversity seen within or between species can be monitored and studied further, potentially providing a basis for inferring evolutionary and phylogenetic relationships between species.

Non-coding and intergenic regions typically evolve at a faster rate than protein-coding genes. Thus, it is expected to see greater variation between species in these non-coding regions. Mitochondrial intergenic spacers have been targeted for use as species-specific markers in other organisms, including COI-COII intergenic region (McFadden *et al.*, 2004; Bandyopadhyay *et al.*, 2008), $tRNA^{leu}$ – COII (Chávez-Galarza *et al.*, 2021), $tRNA^{cys}$ - $tRNA^{asn}$ (Futoshi, 2006), atp6 – COX3 (Andrianov *et al.*, 2010), $tRNA^{ile}$ – $tRNA^{gln}$ and, $tRNA^{gln}$ – $tRNA^{met}$ (da Silva *et al.*, 2009). Since mitochondrial coding regions have had little success in differentiating between the closely related species of interest in this study, the use of non-coding intergenic regions as species-specific markers is the next point of call.

In this study, mitochondrial intergenic region $tRNA^{ile}$ – $tRNA^{gln}$ (denoted as intergenic region I) and $tRNA^{gln}$ – $tRNA^{met}$ (denoted as intergenic region II) were amplified using a single primer pair and investigated for use as species-specific markers for the accurate identification of five economically important tephritid fruit fly species in South Africa.

3.2. Methods and materials

3.2.1. Sample collection

Colony-reared insects and larvae were provided by Citrus Research International (CRI) in Mbombela, Mpumalanga, South Africa, from established colonies (Table S1). Confirmation of the identities of fruit fly species in the colonies (adult specimens from colonies refreshed in the period 2020-2021) was performed by Marc De Meyer, Royal Museum for Central Africa, on 21 February 2022. Wild fruit fly specimens used in this study were collected from traps (Table S2). *Ceratitis* flies were trapped with McPhail type bucket traps baited with enriched ginger root oil (EGO lure) (Insect Science, Tzaneen, South Africa), and *B. dorsalis* flies were trapped with Chempac bucket traps baited with methyl eugenol (ME) (Invader lure, River Bioscience, Gqeberha, South Africa). Fruit fly specimens were stored in 100% ethanol at 4°C until processed.

3.2.2. DNA extraction and species identification

DNA extracts used for high throughput sequencing were obtained from single, adult-male colony insects following an adapted protocol by Sunnucks and Hales (Sunnucks and Hales, 1996), with TNES buffer (50 mM Tris, pH 7.5, 400 mM NaCl, 20 mM EDTA, 0.5 % SDS) substituted for 180 µl ATL buffer (Qiagen). Incubation time was lengthened overnight at 56°C. RNase A was added to the supernatant after NaCl precipitation, and the second precipitation was performed with ice-cold 100% isopropanol overnight at -20°C. DNA concentration and quality were quantified using a NanoDrop 2000 spectrophotometer and a Qubit dsDNA BR assay kit (Invitrogen). Total DNA was also extracted for PCR from colony adults, colony larvae, and wild, trap-collected insects following the destructive protocol of the DNeasy Blood and Tissue Kit (Qiagen), where the whole body of the fruit fly was used.

Each adult colony male specimen in this study underwent molecular identification using universal primer set CI-J2183 and TL2-N3014 for amplification and Sanger sequencing of the COI gene (Simon *et al.*, 1994). The PCR was performed in a total volume of 25 µl containing 1x Kapa Taq buffer A (KAPA Biosystems), 0.2 mM dNTP mix (Thermo Scientific), 0.4 µM of each primer (CI-J2183 and TL2-N3014), and 0.05 U KAPA Taq DNA Polymerase (KAPA Biosystems). The cycling conditions included an initial denaturation step at 94°C for 5 min, followed by 35 cycles of 94°C for 30 s, annealing at 50°C for 30 s and extension at 72°C for 45 s. The final extension was performed at 72°C for 7 min.

All specimens in this study, colony-reared larvae, colony-reared adults, and wild adults underwent identification using a multiplex PCR assay developed for the identification of these five flies following the protocol outlined in the study Andrews *et al* (2022).

3.2.3. High Throughput Sequencing

DNA extracts from two specimen per species were sent for high throughput sequencing at Macrogen (South Korea). Library construction and high throughput sequencing of the colony insects were performed by Macrogen on the Illumina NovaSeq 6000 platform (2 x 150 bp paired-end reads). The TruSeq DNA PCR-Free

Kit was used for library preparation of the samples *C. rosa* 2, *C. quilicii* 1 & 2 and *C. cosyra* 2; and the TruSeq Nano DNA Kit was used for library preparation of samples *C. capitata* 1 & 2, *C. rosa* 1, *C. cosyra* 1, and *B. dorsalis* 1 & 2.

3.2.4. Mitogenome assembly and annotation

Sequencing reads were quality checked using FastQC, all reads and bases were of good quality and further quality checking or trimming was not required. Reference-based assembly was performed with MITObim (Hahn, Bachmann and Chevreux, 2013) using *Ceratitis fasciventris* (GenBank accession NC_035497.1) (Drosopoulou *et al.*, 2017) as a reference template. Assembly was implemented with the following parameters; job=genome, mapping accurate, technology=solexa, parameters=-NW:cmrnl=war, start <1>, end <30>. De novo assembly was performed in CLC genomics workbench version 11.0.1 (Qiagen) using the parameters; automatic bubble size, automatic word size, map reads back to contigs (slow), minimum contig length = 200, mismatch cost = 2, insertion cost = 3, deletion cost = 3, Length fraction = 0.5, and similarity fraction = 0.8. The CLC de novo assembled mitogenome and reference-based assembly for each specimen were aligned, and regions with discrepancies between the two methods were validated with Sanger sequencing (see 3.2.5.). The consensus sequence taken from the alignment of the two assembled mitogenomes and validated variable regions for each specimen was used for manual curation. Manual curation was performed by aligning consensus sequences to relevant reference sequences, *C. capitata* (NC_000857.1) (Spanos *et al.*, 2000), *C. cosyra* (MT036783.1), *C. quilicii*, (MT998948.1), *C. rosa* (MT997010.1) (Zhang *et al.*, 2021) and *B. dorsalis* (KT343905.1), to confirm the starting and ending point of each mitogenome. Mitochondrial genome annotations were performed using the MITOs web server with the parameter “genetic code: 05 – invertebrate” (Bernt *et al.*, 2013) and checked by manually translating the coding domains.

3.2.5. Validation of variable regions

The mitochondrial intergenic region between $tRNA^{ile}$ and $tRNA^{gln}$ was validated using the primer pair trnI_trnQ_F/R. The non-coding region at the 3’end of the genome, known as the control region (CR) and origin of replication (Fig. 3.1.), was validated using primer pair rrnS_trnI_F/R (Table 3.1). Both primer sets were designed with Oligo Explorer 1.1.2 (Gene Link) and synthesised by IDT. PCR reactions were performed in a total volume of 25 µl, with 1x Kapa Taq buffer A (Kapa Biosystems), 0.2 mM dNTP mix (Thermo Scientific), 0.4 µM of each primer (trnI-trnQ _F/R or rrnS_trnI_F/R), 0.05 U KAPA Taq DNA Polymerase (Kapa Biosystems). Cycling consisted of initial denaturation at 94°C for 5 min, with 30 cycles of 94°C for 30 s, annealing at 50°C for 30 s, extension at 72°C for 30 s (primer set trnI_trnQ_F/R), or 72°C for 1 min and 20 s (primer set rrnS_trnI_F/R). The final extension for both primer sets was performed at 72°C for 7 min.

PCR products were visualised on a 2% agarose-TAE gel, purified using the Zymoclean Gel DNA Recovery Kit (Zymo Research), and sent for bidirectional Sanger sequencing at the Central Analytical Facilities at Stellenbosch University.

3.2.6. Primer design

The complete mitochondrial genomes of 30 specimens (Table S3) belonging to the five species of interest and the 10 complete mitochondrial genomes assembled in this study were aligned using CLC Genomics workbench 11.0.1 (Qiagen). Sites with variability between species were visually identified. One primer set (IDT), Mito_F/Mito_R, was designed to amplify mitochondrial DNA between $tRNA^{ile}$ and $tRNA^{met}$ (Fig. 3.1.) using Oligo Explorer 1.1.2 (GeneLink) (Table 3.1).

Table 3.1. Nucleotide sequence, location, and parameters of the primers designed in this study. ON861824 refers to the complete mitochondrial genome of *Bactrocera dorsalis 2*.

Primer	Sequence 5'-3'	Location in ON861824	Tm (°C)	GC%
trnI_trnQ_F	TGAATTGCCTGACAAAAGGG	3-22	53.5	45.0
trnI_trnQ_R	GGTATGAACCCAGTAGCTTA	215-234	51.1	45.0
rrnS_trnI_F	GCTGGCACAAATTTAACCAA	14787-14806	52.0	40.0
rrnS_trnI_R	CCCTTTTGTCAGGCAATTCA	3-22	53.5	45.0
Mito_F	TGACAAAAGGGTTACCTTGATAGGG	12-36	56.6	44.0
Mito_R	ACCCAGTAGCTTAATTAGCTTATCT	203-227	53.4	36.0

3.2.7. High resolution melt analysis

Five adult colony insects, five colony larvae, and 10 wild, trap-collected insects per species were subjected to PCR. PCRs were performed on a Qiagen Rotor-Gene Q thermal cycler. Reactions contained 1X Kapa Taq Buffer A (Kapa Biosystems), 0.4 µM forward primer (IDT), 0.4 µM reverse primer (IDT), 0.2 mM dNTP mix (Thermo Scientific), 1.5 µM SYTO-9 (Invitrogen), 0.05 U KAPA Taq (Kapa Biosystems) and 100 ng DNA. Cycling was conducted on a 36-well carousel with auto gain optimisation performed before the first acquisition, initial hold at 94°C for 5 min, followed by 45 cycles of 94°C for 30 s, annealing at 55°C for 30 s, and extension at 72°C for 30 s. High-resolution melting curves of the PCR amplicons were obtained with temperatures ranging from 70°C to 90°C, with a 0.1°C increase in temperature every two seconds. HRM curve analysis was performed with Rotor-Gene Q software version 2.3.5 (Qiagen). Amplicons were validated through visualisation on a 2% agarose gel to assess specificity (Supplementary Figure S1), and bi-directional Sanger sequencing at the Central Analytical Facilities at Stellenbosch University for downstream sequence analysis.

3.2.8. Sequence analysis

Multiple sequence alignments of the Mito_F/R amplicon were conducted in CLC genomics workbench version 11.0.1 (Qiagen). The alignment consisted of the 40 complete mitochondrial genomes utilised for primer

design and 20 amplicon sequences per species. Amplicons were queried against the NCBI BLASTn database to identify any confounding species with a high sequence similarity to the queried specimens. The amplicons were also queried against a local BLASTn database created on CLC genomics workbench version 11.0.1 (Qiagen), consisting of only the mitochondrial genomes of the five fruit fly species of concern in this study.

3.3. Results

3.3.1. Species identity confirmation

Colony adult males and wild trap-collected specimens underwent morphological identification and identity confirmation using BLASTn analysis querying the COI gene, which was amplified using primer pair CI-J2183 and TL2-N3014 (Simon *et al.*, 1994). All adult specimens used in this study were identified to species level through morphological identification using published keys (Virgilio, White and De Meyer, 2014). COI could differentiate *C. capitata*, *C. cosyra* and *B. dorsalis* to species level. However, the high sequence similarity between *C. rosa*, *C. quilicii*, and the FARQ complex prevented the differentiation of these species. Although *C. capitata* and *Ceratitis caetrata* Munro share high sequence similarity in COI, the latter fly is not present in South Africa (De Meyer, 2000).

3.3.2. Complete mitochondrial genome assembly and annotation

An average of 165 932 970 reads per sample (STD = 4 195 444) were generated with high throughput sequencing (HTS). Mitogenomes assembled with CLC genomics workbench 11.0.1 (Qiagen) had an average read coverage of 7 740 (STD = 2 980.67). Mitogenomes assembled with MITObim (Hahn, Bachmann and Chevreux, 2013) had an average coverage of 7 196 (STD = 2 714.31). The final accessions were generated with consensus sequences from an alignment of the two assembly methods. Variation between the two methods was resolved with Sanger sequencing. The complete mitochondrial genomes were high in similarity to existing references (Table S4). In this study, we assembled 10 complete mitochondrial genomes belonging to five different species in the genera *Ceratitis* and *Bactrocera* (Table 3.2). In total, 37 genes were annotated, including 13 protein-coding genes (PGCs), 22 tRNA's and two rRNA's (Fig. 3.1.). All 10 mitochondrial genomes were highly similar in structural organisation, as previously described (Spanos *et al.*, 2000; Yu *et al.*, 2007; Drosopoulou *et al.*, 2021; Zhang *et al.*, 2021).

Table 3.2. Mitochondrial genome assembly statistics for 10 fruit flies of the genera *Ceratitis* and *Bactrocera*, including NCBI GenBank database accessions.

Specimen	Genome length (bp)	AT%	GC%	Accession
Ceratitis capitata 1	15980	77.409	22.584	ON861815
Ceratitis capitata 2	15981	77.417	22.583	ON861816
Ceratitis cosyra 1	15954	76.194	23.806	ON861817
Ceratitis cosyra 2	15951	76.158	23.842	ON861818
Ceratitis quilicii 1	16020	77.197	22.803	ON861819
Ceratitis quilicii 2	16028	77.396	22.604	ON861820
Ceratitis rosa 1	15998	77.322	22.678	ON861821
Ceratitis rosa 2	15998	77.316	22.684	ON861822
Bactrocera dorsalis 1	15916	73.624	26.376	ON861823
Bactrocera dorsalis 2	15915	73.616	26.384	ON861824

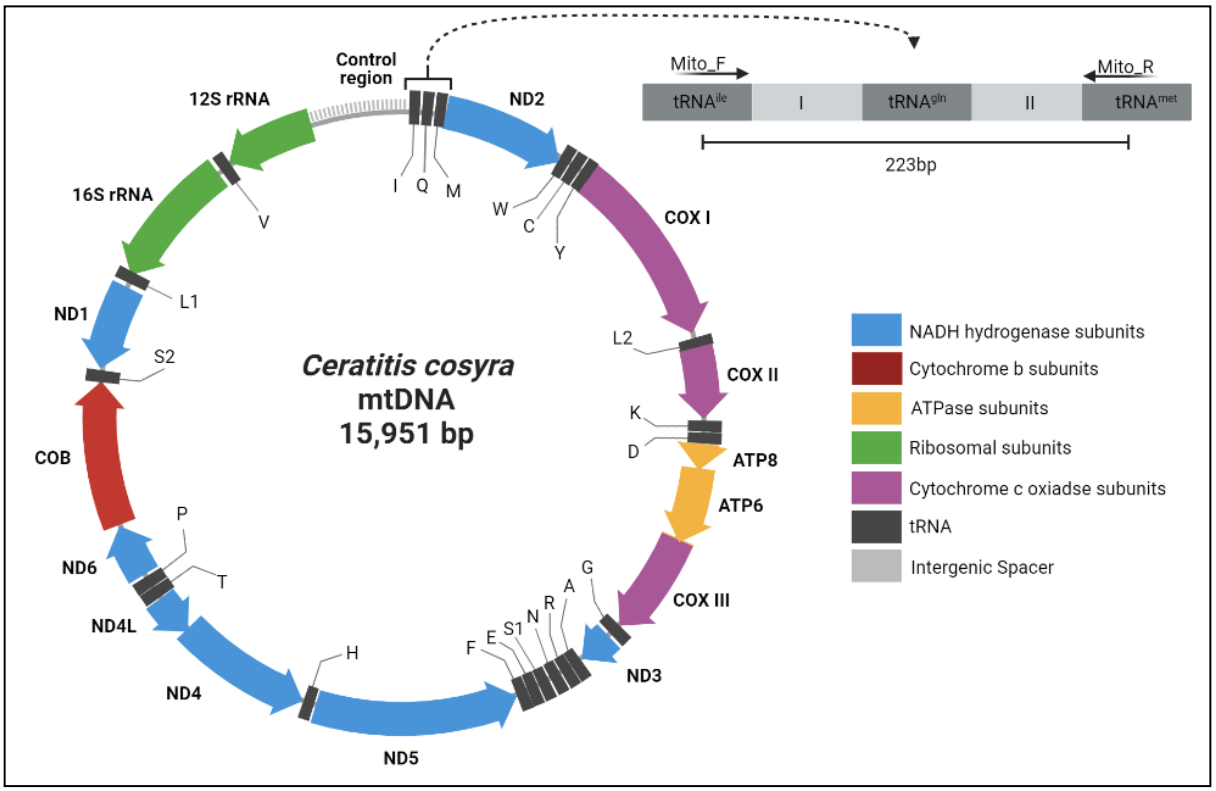


Fig. 3.1. Schematic representation of the complete mitochondrial genome of *Ceratitis cosyra*. The top right corner is an enlarged schematic organisation of the amplicon generated by the primer pair Mito_F/R.

3.3.3. Sequence alignment and primer design

A single primer set, Mito_F/R, was designed to amplify a DNA fragment in all five fruit fly species under investigation (Table 3.1). Each PCR amplicon contained a partial sequence of tRNAile (56 bp), the complete sequence of tRNAgln (69 bp), a partial sequence of tRNAmet (29 bp), and two intergenic spacers (Fig. 3.1.).

The sense primer is located between 12 – 36 bp (tRNAile), and the antisense primer is located between 203 – 227 bp (tRNAmet). Location refers to the Bactrocera dorsalis accession ON861824.High-resolution melt analysis and PCR amplification

The primer set Mito_F/R used in this study was initially designed for use in a high-resolution melt (HRM) for species differentiation based on melting point analysis of the resulting amplicon (Fig. 3.2). The average melt temperature (Tm) and standard deviation of each species-specific amplicon are presented in Table 3.3. However, this approach was unsuccessful due to the low GC content of the selected region in the mitochondrial genome affecting the consistency of dye intercalation. Hence, melt-point intervals were inconsistent within species, and confidence intervals could not be accurately determined. Primer sets with a higher GC content were tested on the same samples as primer set Mito_HRM_F and Mito_HRM_R. These primers were developed in Chapter 4 (Opsin4_capitata, Opsin4_cosyra, Opsin4_quilicii, and Opsin4_rosa). HRM analysis was performed with the same parameters used for the primer set Mito_HRM F & R. The higher GC content in amplicons generated by the opsin4 primer sets correlated to smoother, more consistent amplification plots with stable melting temperatures (Table 3.4) (Fig. 3.3).

Table 3.3. Average melt temperature (Tm) and standard deviation of each fruit fly species-specific amplicon based on high-resolution melt analysis of amplicon Mito_F/R.

Species	Average Tm (°C)	Standard deviation
C. capitata	72.46	0.63
C. cosyra	72.77	0.18
C. quilicii	71.99	0.45
C. rosa	72.06	0.31
B. dorsalis	73.98	0.39

Table 3.4. Comparison of high-resolution melt curves between primer sets with differing GC contents.

Primer set	Species	Amplicon length (bp)	GC content (%)	Fluorescence[1]
Mito_F/R	*C. capitata*	236	17.37	15
	C. cosyra	169	19.53	15-20
	C. quilicii	225	16.00	5
	C. rosa	196	17.86	1-5
	B. dorsalis	206	21.84	65
Opsin4_capitata F & R	*C. capitata*	326	45.40	100
Opsin4_cosyra F&R	*C. cosyra*	182	28.02	35 - 40
Opsin4_quilicii F&R	*C. quilicii*	124	33.06	55 - 65
Opsin4_rosa F&R	*C. rosa*	248	28.63	75 - 80

[1] Fluorescent units as determined by the Rotor-Gene Q software version 2.3.5 (Qiagen)

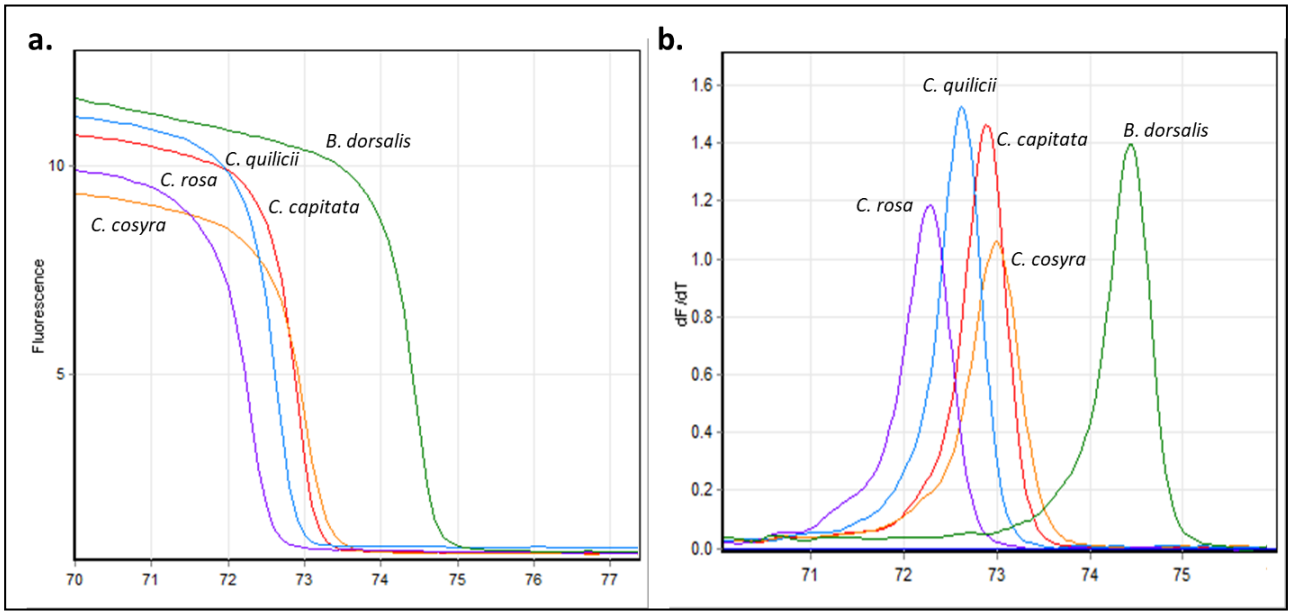


Fig. 3.2. High resolution melting curve analysis for the detection of fruit fly species using Syto-9. Normalised HRM curve (**a.**) and derivative HRM curve (**b.**) for identification with primer pair Mito_F/R.

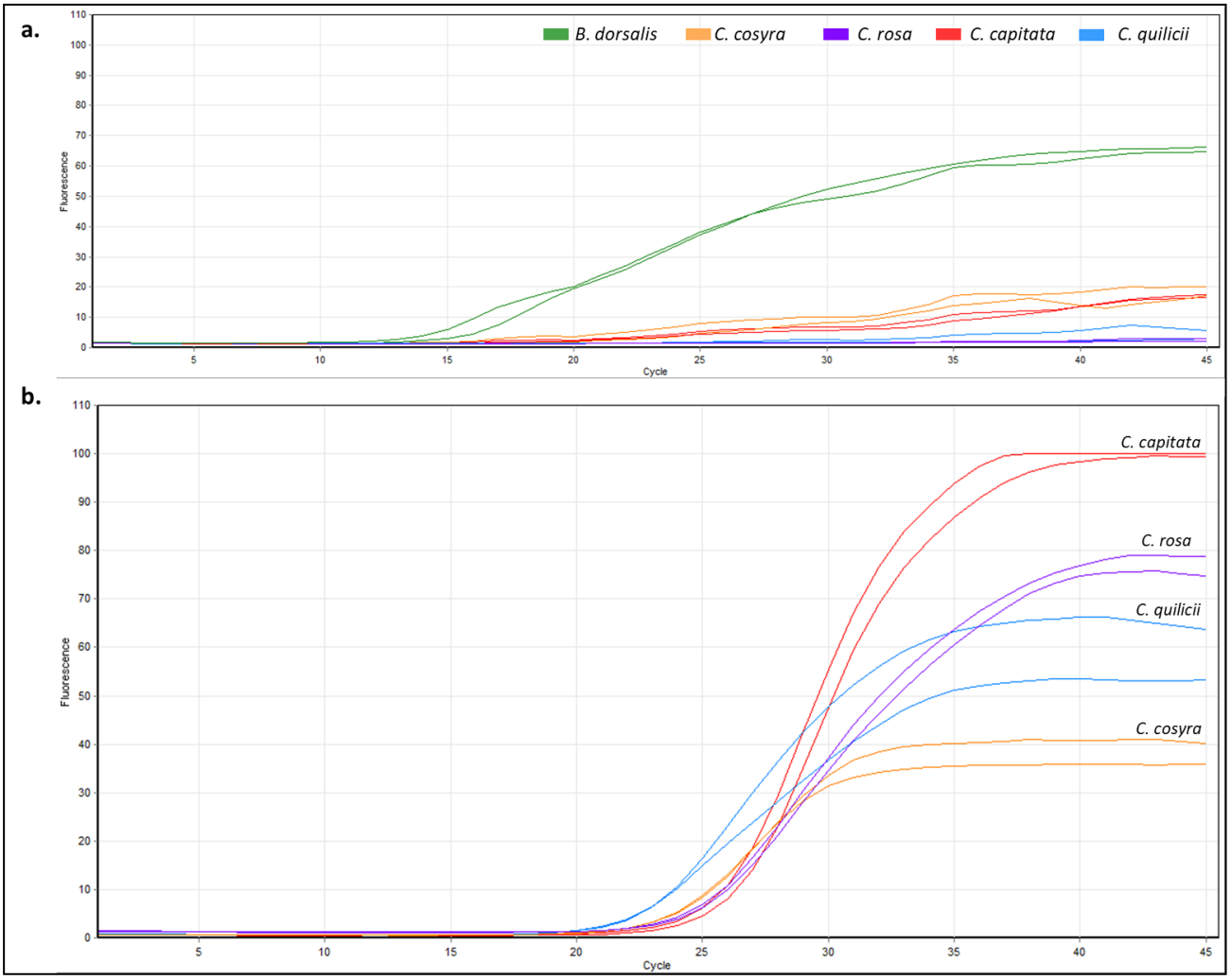


Fig. 3.3. Comparison of amplification curves of (**a.**) amplicons of lower GC content generated with the primer set Mito_F/R, and (**b.**) amplicons of higher GC content generated with Opsin4 primers (Chapter 4). Both curves were obtained using Syto-9 for the detection of fruit fly species.

3.3.4. Sequence analysis

A total of 20 specimens per species underwent PCR and amplicon sequencing. Multiple sequence alignments of the amplicons and available reference sequences demonstrate the ability to differentiate between the five fruit fly species based on intergenic regions (Fig. 3.4.). BLASTn analysis of the whole Mito_F/R amplicon against the NCBI GenBank database highlighted high sequence similarity between *C. quilicii*, *C. fasciventris* (100 %), and *C. anonae* (97.78%), between *C. cosyra*, *Ceratitis pallidula* De Meyer, Mwatawala & Virgilio (97.04%), and *Ceratitis quinaria* (Bezzi) (96.45%), between *B. dorsalis, Bactrocera invadens* Drew, Tsuruta & White (99.51%), *Bactrocera carambolae* Drew & Hancock (99.51%), *Bactrocera philippinensis* Drew & Hancock (99.51%), *Bactrocera papayae* Drew & Hancock (99.51%), *Bactrocera ruiliensis* Wang, Long & Zhang (97.57%), and *Bactrocera thailandica* Drew and Romig (97.57%). Local BLASTn analyses against only the five fruit flies investigated identified each amplicon to species level accurately.

Fig. 3.4. Nucleotide sequence comparison of intergenic regions I (**a.**) and II (**b.**) showing the variation present within each species and the difference in the size of intergenic regions between each species. Dashes represent a gap or missing nucleotide, and dots represent a matching nucleotide. The number of specimens with a particular intergenic sequence is indicated in round brackets next to the species name, and the size of the intergenic region is indicated in square brackets at the end of each nucleotide sequence.

3.4. Discussion

The generation of complete mitochondrial genome sequences adds value to publicly available online databases by providing increasingly extensive resources. Complete mitochondrial genomes provide useful molecular markers for both taxonomic and molecular studies. These markers have been widely used in the study of insects and Tephritidae in particular (Barr *et al.*, 2006, 2012; Dhami *et al.*, 2016; Jiang *et al.*, 2016). The availability and abundance of complete mitochondrial genomes are essential for studying closely related and cryptic species, specifically when existing species identification tools are limited in efficacy.

While the HRM analysis was unable to consistently differentiate between species, the nucleotide sequences generated from these amplicons were applicable in the species identification of the five fruit flies. The variation in size of the complete mitochondrial genome between members of the family Tephritidae is mainly due to variation in non-coding regions and intergenic spacers (Yong *et al.*, 2016). Mitochondrial intergenic regions, including intergenic region I between $tRNA^{ile}$ and $tRNA^{gln}$, have previously been utilised in analyses of phylogeny and genetic distance (da Silva *et al.*, 2009). The length of intergenic region I is a potential tool for differentiation of these five species, *C. capitata* (40 bp), *C. cosyra* (2 bp), *C. quilicii* (62-64 bp), *C. rosa* (34 bp), and B. dorsalis (0 bp) (Fig. 3.4.). The absence of intergenic region I in *B. dorsalis* has previously been described and is common among members of the genus *Bactrocera* (Yu *et al.*, 2007; Yong *et al.*, 2016; Drosopoulou *et al.*, 2019). Due to high similarity, intergenic region II could not differentiate between *C. rosa* and *C. quilicii* (Fig. 3.4.). Furthermore, this region showed greater variation between individuals of the same species than intergenic region I, specifically in the case of *C. capitata*.

BLASTn analysis of amplicons generated with Mito_F/R against a local database consisting of the mitochondrial genomes of only the five flies investigated in this study accurately differentiates and identifies these flies to species level. In comparison, BLASTn analysis of the same amplicons against the NCBI database revealed several potential confounding species. None of the species identified are currently present or have ever been reported in South Africa (De Meyer, 2001; Copeland *et al.*, 2006; Vargas, Piñero and Leblanc, 2015), with the exception being *C. quinaria*. However, this fly is not regarded as a pest of economic importance in commercial fruit in South Africa. Three fruit fly species confounding the BLAST analysis of the *B. dorsalis* Mito_F/R amplicon have been synonymised as *B. dorsalis*, namely, *B. invadens*, *B. philippinensis*, and *B. papayae* (Schutze *et al.*, 2015). The *C. quilicii* Mito_F/R amplicon is highly similar to *C. fasciventris* (96.90 – 100%) and *C. anonae* (93.81 – 97.87%). However, the amplicon does not share a high similarity with that of *C. rosa* (84.44%). This allows for unambiguous differentiation of the cryptic species *C. quilicii* and *C. rosa*, which cannot be differentiated based on COI due to high sequence similarity within the *Ceratitis* FARQ complex (Virgilio *et al.*, 2008; De Meyer *et al.*, 2016; Zhang *et al.*, 2021).

Interestingly, *C. rosa* did not share a high similarity with the other members of the FARQ complex within the Mito_F/R amplicon region. High similarity of the amplicon between *C. quilicii*, *C. fasciventris*, and *C. anonae*

corresponds with previous studies where evidence of gene flow between *C. quilicii* and *C. fasciventris* suggests an admixture event between these two species (Virgilio *et al.*, 2013; Zhang *et al.*, 2021). In addition, it has been reported that *C. fasciventris* and *C. quilicii* or *C. rosa*, which were previously recognised as a single species (De Meyer *et al.*, 2016), can reproduce under laboratory conditions (Erbout, De Meyer and Lens, 2008).

Variation in noncoding regions is not uncommon as intergenic spacers can differ considerably, even in the case of closely related species. Notably in this study, the size of the intergenic region I is relatively consistent within species and significantly different between species to allow for unambiguous differentiation of these five fruit fly species. Species identification based on the size of mitochondrial intergenic regions has shown to be successful in a wide range of bacteria and is emerging in insect phylogenetics (McClelland, Petersen and Welsh, 1992; Baele *et al.*, 2000; da Silva *et al.*, 2009; Du *et al.*, 2017; Xu, Zhou and Wan, 2022). It is important to frame this work in a greater context of mitochondrial datasets being at the forefront of the exploration of evolutionary relationships. An advantage of the identification process described is that the primer set used is universal as the target regions are located within conserved tRNA's. In principle, this tool has the potential to study multiple genera within the family Tephritidae.

This study reports 10 complete mitochondrial genomes for the fruit flies *C. capitata*, *C. cosyra*, *C. rosa*, *C. quilicii* and *B. dorsalis*. The availability of these mitogenomes will aid future studies regarding tephritid fruit flies. We propose that the length of intergenic region I between $tRNA^{ile}$ and $tRNA^{gln}$, and multiple sequence alignment of the amplicon Mito_F/R against reference sequences can be used as informative species-specific markers for differentiation of these five tephritid flies present in South Africa. The identification tool described in this study can be used as an alternative to traditional DNA barcoding for accurate species identification of these flies for routine pest monitoring practices in South Africa. Furthermore, the inherent impact of studying intergenic spacers such as those described in this study offers advancement and utility in global fruit fly research and diversity studies and can potentially be expanded for phylogenetic and taxonomic evaluation.

3.5. References

Andrews, K. J. *et al.* (2022) 'A multiplex PCR assay for the identification of fruit flies (Diptera: Tephritidae) of economic importance in South Africa', *Scientific Reports*, 12(1), p. 13089. doi: 10.1038/s41598-022-17382-x.

Andrianov, B. *et al.* (2010) 'Comparative analysis of the mitochondrial genomes in *Drosophila virilis* species group (Diptera: Drosophilidae)', *Trends in Evolutionary Biology*, 2(1), p. 4. doi: 10.4081/eb.2010.e4.

Baele, M. *et al.* (2000) 'Application of tRNA intergenic spacer PCR for identification of *Enterococcus* species', *Journal of Clinical Microbiology*, 38(11), pp. 4201–4207.

Balmès, V. and Mouttet, R. (2017) 'Development and validation of a simplified morphological identification key for larvae of tephritid species most commonly intercepted at import in Europe', *EPPO Bulletin*, 47(1), pp. 91–99. doi: 10.1111/epp.12369.

Bandyopadhyay, P. K. *et al.* (2008) 'The mitochondrial genome of *Conus textile*, *coxI–coxII* intergenic sequences and Conoidean evolution', *Molecular Phylogenetics and Evolution*, 46(1), pp. 215–223. doi: 10.1016/j.ympev.2007.08.002.

Barr, N. *et al.* (2006) 'Molecular diagnostics of economically important *Ceratitis* fruit fly species (Diptera: Tephritidae) in Africa using PCR and RFLP analyses', *Bulletin of Entomological Research*, 96, pp. 505–521. doi: 10.1079/BER2006452.

Barr, N. B. *et al.* (2012) 'Molecular Identification of *Ceratitis capitata* (Diptera: Tephritidae) Using DNA Sequences of the COI Barcode Region', *Annals of the Entomological Society of America*, 105(2), pp. 339–350. doi: 10.1603/AN11100.

Bernt, M. *et al.* (2013) 'MITOS: Improved de novo metazoan mitochondrial genome annotation', *Molecular Phylogenetics and Evolution*, 69(2), pp. 313–319. doi: 10.1016/j.ympev.2012.08.023.

Boykin, L. M. *et al.* (2012) 'Species Delimitation and Global Biosecurity', *Evolutionary Bioinformatics*, 8, p. EBO.S8532. doi: 10.4137/EBO.S8532.

Bragard, C. *et al.* (2020) 'Pest categorisation of non-EU Tephritidae', *EFSA Journal*, 18(1). doi: 10.2903/j.efsa.2020.5931.

Chávez-Galarza, J. *et al.* (2021) 'Mitochondrial DNA Variation in Peruvian Honey Bee (*Apis mellifera* L.) Populations Using the tRNAleu-cox2 Intergenic Region', *Insects*, 12(7), p. 641. doi: 10.3390/insects12070641.

Copeland, R. S. *et al.* (2006) 'Geographic Distribution, Host Fruit, and Parasitoids of African Fruit Fly Pests *Ceratitis anonae*, *Ceratitis cosyra*, *Ceratitis fasciventris*, and *Ceratitis rosa* (Diptera: Tephritidae) in Kenya', *Annals of the Entomological Society of America*, 99(2), pp. 261–278.

Danecek, P. *et al.* (2021) 'Twelve years of SAMtools and BCFtools', *GigaScience*, 10(2). doi: 10.1093/gigascience/giab008.

da Silva, N. M. *et al.* (2009) 'Characterization of mitochondrial control region, two intergenic spacers and tRNAs of *Zaprionus indianus* (Diptera: Drosophilidae)', *Genetica*, 137(3), pp. 325–332. doi: 10.1007/s10709-009-9396-5.

De Meyer, M. (2000) 'Systematic revision of the subgenus *Ceratitis* MacLeay s.s. (diptera, Tephritidae)', *Zoological Journal of the Linnean Society*, 128, pp. 439–467. doi: 10.1006/zils. 1999.01 98.

De Meyer, M. (2001) 'Distribution patterns and host-plant relationships within the genus *Ceratitis* MacLeay (Diptera: Tephritidae) in Africa', *Cimbebasia*, 17, pp. 219–228.

De Meyer, M. *et al.* (2016) 'Description of new *Ceratitis* species (Diptera: Tephritidae) from Africa, or how morphological and DNA data are complementary in discovering unknown species and matching sexes', *European Journal of Taxonomy*, 233(233), pp. 1–23. doi: 10.5852/ejt.2016.233.

Dhami, M. K. *et al.* (2016) 'A real-time PCR toolbox for accurate identification of invasive fruit fly species', *Journal of Applied Entomology*, 140(7), pp. 536–552. doi: 10.1111/jen.12286.

Drosopoulou, E. *et al.* (2017) 'The chromosomes and the mitogenome of *Ceratitis fasciventris* (Diptera: Tephritidae): two genetic approaches towards the *Ceratitis* FAR species complex resolution', *Scientific Reports*, 7(1), p. 4877. doi: 10.1038/s41598-017-05132-3.

Drosopoulou, E. *et al.* (2019) 'The Complete Mitochondrial Genome of *Bactrocera carambolae* (Diptera: Tephritidae): Genome Description and Phylogenetic Implications', *Insects*, 10(12), p. 429. doi: 10.3390/insects10120429.

Drosopoulou, E. *et al.* (2021) 'The complete mitochondrial genomes of *Ceratitis rosa* and *Ceratitis quilicii* , members of the *Ceratitis* FAR species complex (Diptera: Tephritidae)', *Mitochondrial DNA Part B*, 6(3), pp. 1039–1041. doi: 10.1080/23802359.2021.1899073.

Du, C. *et al.* (2017) 'Mitochondrial genomes of blister beetles (Coleoptera, Meloidae) and two large intergenic spacers in *Hycleus* genera', *BMC Genomics*, 18(1), p. 698. doi: 10.1186/s12864-017-4102-y.

Ekesi, S. and Billah, M. K. (eds) (2006) *Field guide to the management of economically important tephritid fruit flies in Africa*. 2nd edn. Nairobi, Kenya: ICIPE Science Press.

Erbout, N., De Meyer, M. and Lens, L. (2008) 'Hybridization between two polyphagous fruit-fly species (Diptera: Tephritidae) causes sex-biased reduction in developmental stability', *Biological Journal of the Linnean Society*, (93), pp. 579–588.

Frey, J. E. *et al.* (2013) 'Developing diagnostic SNP panels for the identification of true fruit flies (Diptera: Tephritidae) within the limits of COI-based species delimitation', *BMC Evolutionary Biology*, 13(1), p. 106. doi: 10.1186/1471-2148-13-106.

Futoshi, A. (2006) 'A nocel mitochondrial intergenic spacer reflecting population structure of the Pacific oyster', *Jounral of Applied Genetics*, 47(2), pp. 119–123.

Grout, T. G. (2015) 'The Status of Citrus IPM in South Africa', *Acta Horticulturae*. Edited by B. Sabater-Muñoz, (1065), pp. 1091–1095. doi: 10.17660/ActaHortic.2015.1065.137.

Hahn, C., Bachmann, L. and Chevreux, B. (2013) 'Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach', *Nucleic Acids Research*, 41(13), pp. e129–e129. doi: 10.1093/nar/gkt371.

IPPC (2021) 'ISPM20 Guidelines for a phytosanitary import regulatory system'. Available at: https://www.ippc.int/en/publications/602/.

Jiang, F. *et al.* (2014) 'Existence of species complex largely reduced barcoding success for invasive species of Tephritidae: a case study in *Bactrocera* spp.', *Molecular Ecology Resources*, 14(6), pp. 1114–1128. doi: 10.1111/1755-0998.12259.

Jiang, F. *et al.* (2016) 'A high-throughput detection method for invasive fruit fly (Diptera: Tephritidae) species based on microfluidic dynamic array', *Molecular Ecology Resources*, 16(6), pp. 1378–1388. doi: 10.1111/1755-0998.12542.

Li, H. and Durbin, R. (2009) 'Fast and accurate short read alignment with Burrows-Wheeler transform.', *Bioinformatics (Oxford, England)*, 25(14), pp. 1754–60. doi: 10.1093/bioinformatics/btp324.

Louzeiro, L. R. F. *et al.* (2021) 'Incidence of frugivorous flies (Tephritidae and Lonchaeidae), fruit losses and the dispersal of flies through the transportation of fresh fruit', *Journal of Asia-Pacific Entomology*, 24(1), pp. 50–60. doi: 10.1016/j.aspen.2020.11.006.

Malacrida, A. R. *et al.* (2007) 'Globalization and fruitfly invasion and expansion: the medfly paradigm', *Genetica*, 131(1), pp. 1–9. doi: 10.1007/s10709-006-9117-2.

Manrakhan, A. (2020) 'Fruit Fly', in Grout, T. (ed.) *Integrated Production Guidelines for export citrus.* Nelspruit, South Africa: Citrus Research International, pp. 1–10.

Manrakhan, A., Venter, J. H. and Hattingh, V. (2015) 'The progressive invasion of *Bactrocera dorsalis* (Diptera: Tephritidae) in South Africa', *Biological Invasions*, 17(10), pp. 2803–2809. doi: 10.1007/s10530-015-0923-2.

McClelland, M., Petersen, C. and Welsh, J. (1992) 'Length polymorphisms in tRNA intergenic spacers detected by using the polymerase chain reaction can distinguish Streptococcal strains and species', *Journal of Clinical Microbiology*, 30(6), pp. 1499–1504.

McFadden, C. S. *et al.* (2004) 'Variation in Coding (NADH Dehydrogenase Subunits 2, 3, and 6) and Noncoding Intergenic Spacer Regions of the Mitochondrial Genome in *Octocorallia* (Cnidaria: Anthozoa)', *Marine Biotechnology*, 6(6), pp. 516–526. doi: 10.1007/s10126-002-0102-1.

Mommsen, W. and Bester, H. (2018) 'CRI IPM & Disease Management Workshops on Raising the Standards of Citrus production in South Africa', *Citrus Research International*, pp. 102–106.

Prinsloo, G. L. and Uys, V. M. (eds) (2015) *Insects of Cultivated Plants and Natural Pastures in Southern Africa*. Entomological Society of Southern Africa.

Qin, Y. *et al.* (2018) 'Population structure of a global agricultural invasive pest, *Bactrocera dorsalis* (Diptera: Tephritidae)', *Evolutionary Applications*, 11(10), pp. 1990–2003. doi: 10.1111/eva.12701.

Richardson, M. F. *et al.* (2012) 'Population Genomics of the *Wolbachia* Endosymbiont in *Drosophila melanogaster*', *PLoS Genetics*. Edited by A. Kopp, 8(12), p. e1003129. doi: 10.1371/journal.pgen.1003129.

Schutze, M. K. *et al.* (2015) 'Synonymization of key pest species within the *Bactrocera dorsalis* species complex (Diptera: Tephritidae): taxonomic changes based on a review of 20 years of integrative morphological, molecular, cytogenetic, behavioural and chemoecological data', *Systematic Entomology*, 40(2), pp. 456–471. doi: 10.1111/syen.12113.

Signor, S. (2017) 'Population genomics of *Wolbachia* and mtDNA in *Drosophila simulans* from California', *Scientific Reports*, 7(1), p. 13369. doi: 10.1038/s41598-017-13901-3.

Simon, C. *et al.* (1994) 'Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers', *Annals of the Entomological Society of America*, 87(6), pp. 651–701. doi: 10.1093/aesa/87.6.651.

Spanos, L. *et al.* (2000) 'The mitochondrial genome of the Mediterranean fruit fly, *Ceratitis capitata*.', *Insect Molecular Biology*, 9(2), pp. 139–144.

Steck, G. . (2004) 'Fruit Flies (Diptera: Tephritidae)', in *Encyclopedia of Entomology*. Dordrecht: Springer. doi: 10.1007/0-306-48380-7_1706.

Sunnucks, P. and Hales, D. F. (1996) 'Numerous Transposed Sequences of Mitochondrial Cytochrome Oxidase I-II in Aphids of the Genus *Sitobion* (Hemiptera: Aphididae)', *Molecular biology and evolution*, 13(3), pp. 510–524. doi: 10.1093/oxfordjournals.molbev.a025612.

Vargas, R., Piñero, J. and Leblanc, L. (2015) 'An Overview of Pest Species of *Bactrocera* Fruit Flies (Diptera: Tephritidae) and the Integration of Biopesticides with Other Biological Approaches for Their Management with a Focus on the Pacific Region', *Insects*, 6(2), pp. 297–318. doi: 10.3390/insects6020297.

Venter, J. H., Baard, C. W. L. and Barnes, B. N. (2021) 'Area-wide management of Mediterranean fruit fly with the sterile insect technique in South Africa: new production and management techniques pay dividends', in *Area-Wide Integrated Pest Management*. CRC Press, pp. 129–141.

Virgilio, M. *et al.* (2008) 'Molecular evaluation of nominal species in the *Ceratitis fasciventris*, *C. anonae*, *C. rosa* complex (Diptera: Tephritidae)', *Molecular Phylogenetics and Evolution*, 48(1), pp. 270–280. doi: 10.1016/j.ympev.2008.04.018.

Virgilio, M. *et al.* (2012) 'Identifying Insects with Incomplete DNA Barcode Libraries, African Fruit Flies (Diptera: Tephritidae) as a Test Case', *PLoS ONE*. Edited by D. Steinke, 7(2), p. e31581. doi: 10.1371/journal.pone.0031581.

Virgilio, M. *et al.* (2013) 'Cryptic diversity and gene flow among three African agricultural pests: *Ceratitis rosa* , *Ceratitis fasciventris* and *Ceratitis anonae* (Diptera, Tephritidae)', *Molecular Ecology*, 22(9), pp. 2526–2539. doi: 10.1111/mec.12278.

Virgilio, M. *et al.* (2019) 'An integrated diagnostic setup for the morphological and molecular identification of the *Ceratitis* FAR complex (*C. anonae, C. fasciventris, C. rosa , C. quilicii* , Diptera, Tephritidae)', *Bulletin of Entomological Research*, 109(3), pp. 376–382. doi: 10.1017/S0007485318000615.

Virgilio, M., White, I. and De Meyer, M. (2014) 'A set of multi-entry identification keys to African frugivorous flies (Diptera, Tephritidae)', *ZooKeys*, 428, pp. 97–108. doi: 10.3897/zookeys.428.7366.

Waugh, J. (2007) 'DNA barcoding in animal species: progress, potential and pitfalls', *BioEssays*, 29(2), pp. 188–197. doi: 10.1002/bies.20529.

Whatson, M. (2020) 'Decision To Revise Import Requirements for the Importation of Fresh Citrus From South Africa Into the United States', *Federal Registar*, 85(215).

Xu, M., Zhou, S. and Wan, X. (2022) 'Phylogenetic Implication of Large Intergenic Spacers: Insights from a Mitogenomic Comparison of *Prosopocoilus* Stag Beetles (Coleoptera: Lucanidae)', *Animals*, 12(13), p. 1595. doi: 10.3390/ani12131595.

Yong, H.-S. *et al.* (2016) 'Complete Mitochondrial Genome of Three *Bactrocera* Fruit Flies of Subgenus *Bactrocera* (Diptera: Tephritidae) and Their Phylogenetic Implications', *PLOS ONE*. Edited by B.-S. Yue, 11(2), p. e0148201. doi: 10.1371/journal.pone.0148201.

Yu, D. J. *et al.* (2007) 'The complete nucleotide sequence of the mitochondrial genome of the oriental fruit fly, *Bactrocera dorsalis* (Diptera: Tephritidae)', *Gene*, 396(1), pp. 66–74. doi: 10.1016/j.gene.2007.02.023.

Zhang, Y. *et al.* (2021) 'Phylogenomic resolution of the *Ceratitis* FARQ complex (Diptera: Tephritidae)', *Molecular Phylogenetics and Evolution*, 161, p. 107160. doi: 10.1016/j.ympev.2021.107160

CHAPTER 4. A MULTIPLEX PCR ASSAY FOR THE IDENTIFICATION OF FRUIT FLIES (DIPTERA: TEPHRITIDAE) OF ECONOMIC IMPORTANCE IN SOUTH AFRICA

4.1. Background

Tephritidae is an agriculturally important family with many fruit fly species known to cause extensive damage to commercial fruit (White and Elson-Harris, 1992). Quarantine restrictions are in place to limit any further spread of these fruit fly pests. In South Africa, five economically important fruit flies are present that can potentially affect the production and export of commercial fruit (De Villiers *et al.*, 2013; Manrakhan, Venter and Hattingh, 2015; De Meyer *et al.*, 2016). They are *Ceratitis capitata* (Wiedemann), Mediterranean fruit fly; *Ceratitis cosyra* (Walker), marula fly; *Ceratitis rosa* (Karsch), Natal fly*; Ceratitis quilicii* De Meyer, Mwatawala & Virgilio, Cape fly; and *Bactrocera dorsalis* (Hendel), the Oriental fruit fly. *Ceratitis quilicii* is a recently described species (De Meyer *et al.*, 2016), hence the pest status and host range of this species in commercial fruit production areas in South Africa are still being determined. The *Ceratitis* species are of Afrotropical origin (De Meyer, 2001) while *B. dorsalis* is of Asian origin and was introduced in the northeastern parts of South Africa in 2013 (Manrakhan, Venter and Hattingh, 2015). The five fruit fly species are polyphagous (attacking fruit from different plant families) (Virgilio, White and De Meyer, 2014), and two of them, *C. capitata* and *B. dorsalis*, have demonstrated a high affinity for invasiveness with significant expansion of their distribution beyond their native ranges (Malacrida *et al.*, 2007; Qin *et al.*, 2018). This is a major challenge for horticultural and export industries, particularly with the increasing frequency of international trade (Louzeiro *et al.*, 2021). These five fruit flies are currently the only major tephritid pests of commercial fresh fruit produced primarily for export from South Africa.

South Africa is a significant producer and exporter of fresh fruit. In the 2019/2020 production season, over 6.5 million metric tons of fruit were produced, and more than half of the total produce was exported (Fruit South Africa, *2020 Key Fruit Statistics*). Fruit fly pests are of phytosanitary significance for fresh fruit exported from South Africa. It is not uncommon for multiple fruit fly species to infest the same commercial fruit, as the host range of these fruit flies often overlap (Rasolofoarivao, Raveloson Ravaomanarivo and Delatte, 2022). South African fruit must meet the country-specific phytosanitary requirements of the export markets to prevent the entry of fruit fly pests (EPPO, 2020). The interception of phytosanitary pests on consignments at Ports of Entry (PoE) can result in the destruction of the commodity or return of the commodity to the country of origin (Whatson, 2020). The time required to accurately identify any pests present in consignments delays the shipment of fresh produce. Fresh fruit and vegetables may be detained for days while undergoing inspection, reducing their economic value. The European Union (EU), an important export market for fresh fruit from South Africa, has zero-tolerance enforcement for non-EU Tephritidae, including all fruit fly pests in South Africa, except *C. capitata*, which is an established pest in the EU (Bragard *et al.*, 2020). There is a need to be able to distinguish between *C. capitata* and the other four fruit fly pests during inspection before and

after export for markets such as the EU. Therefore, a rapid method to accurately identify the five fruit fly pests infesting fresh fruit in South Africa is essential.

Inspection and survey of fruit fly species are often largely reliant on morphological identification of specimens by expert taxonomists and published keys (White and Elson-Harris, 1992; Virgilio, White and De Meyer, 2014; Balmès and Mouttet, 2017). The morphological identification of fruit flies to species level can be more reliably made at the adult stage, either emerged adults from infested fruit or adults collected from traps, using these keys. The difficulty arises in differentiating between cryptic species or damaged adult specimens where few distinguishing morphological differences exist, and female specimens appear near identical. However, when immature stages are intercepted, either eggs or larvae in fruit or pupae in soil, and development to adulthood is not practical due to time sensitivity, identification to species level using molecular methods would be more appropriate (Boykin *et al.*, 2012). There are keys to differentiate between species at the larval stages. Still, these can only be used at the third instar stage and remain problematic if either the specimen is not in good condition, earlier life stages are intercepted, or pupae are found. Larval identification using these keys does not enable the identification of closely related species or species complexes (Balmès and Mouttet, 2017). *Ceratitis rosa* and *C. quilicii* are examples of cryptic species that were previously thought to be the same species until their recent separation in 2016 based on morphological and genetic differences (De Meyer *et al.*, 2016; Virgilio *et al.*, 2019). Research has demonstrated that the taxonomic classification of many of these cryptic fruit fly species cannot reliably be resolved through morphological characterisation alone, where population-level variation can be easily confused with species-level variation (Tan *et al.*, 2013). A molecular-based identification assay may alleviate some of the difficulties experienced in the morphological identification of such closely related species.

Molecular identification tools can offer the advantage of a faster turnaround time as the life stage of the specimen is not a limiting factor. Broad detection assays for tephritid fruit flies have been developed, although they do not reliably allow for identification to species level (Jiang *et al.*, 2016, 2018). Microsatellite markers have been considered for identification of closely related fruit fly species, however, this can be expensive and time consuming where six to 16 markers have to be used, and unambiguous species identification is not possible without prior morphological identification (Delatte *et al.*, 2013, 2014; Virgilio *et al.*, 2013). Molecular identification techniques for fruit flies have been primarily centered around DNA barcoding using cytochrome c oxidase subunit I (COI). Although this method can differentiate between many fruit fly species, it cannot accurately differentiate between species complexes such as FARQ (*C. fasciventris* (Bezzi), *C. anonae* Graham, *C. rosa,* and *C. quilicii)* (Virgilio *et al.*, 2008; De Meyer *et al.*, 2016; Zhang *et al.*, 2021) and the *Bactrocera dorsalis* complex (Jiang *et al.*, 2014). Literature suggests that the failure of COI to identify closely related species may be due to incomplete lineage sorting within these species complexes (Jiang *et al.*, 2014). Misidentifications can be reduced by introducing a distance threshold, where a query sequence is considered unidentifiable if the closest DNA barcode match exceeds the value of the distance

threshold set. However, if the distance threshold is too restrictive, it is at the sacrifice of reduced barcoding accuracy with a higher proportion of discarded queries (Virgilio *et al.*, 2012). DNA barcoding relies on time-consuming DNA sequencing, an additional expenditure not applicable for routine analysis (Van Houdt *et al.*, 2010; Barr *et al.*, 2012). COI has also been used for qPCR and real-time PCR identification assays (Dhami *et al.*, 2016; Jiang *et al.*, 2016). Expansions into other mitochondrial genes for PCR-RFLP analysis and Tephritidae identification have also been explored (Barr *et al.*, 2006, 2012). However, most molecular identification assays based on mitochondrial genes had limitations in identifying closely related species and species complexes. Endosymbiotic bacteria, such as *Wolbachia*, have been identified as potential mechanisms of speciation through cytoplasmic incompatibility (CI) brought about by an infected male mating with an uninfected female or a female infected with a different *Wolbachia* strain which ultimately results in sexual incompatibility producing non-viable embryos (Weeks, Reynolds and Hoffmann, 2002). An example of this phenomenon has been observed in natural populations of the tephritid flies *Rhagoletis cerasi* Loew and *Bactrocera dorsalis*, where fruit fly populations separated by geographical barriers harbour slightly different *Wolbachia* strains (Sun, Cui and Li, 2007; Riegler and Stauffer, 2008). Similarly, studies on the *Anastrepha fraterculus* (Wiedemann) species complex report that the different morphospecies in this complex have distinct *Wolbachia* strains (Prezotto *et al.*, 2017). From these observations, there is potential utility in indexing *Wolbachia* present in different fruit fly species and using this information to identify closely related tephritid species to species level based on their composition of *Wolbachia* infections.

An investigation into the use of genomic regions as opposed to mitochondrial research for tephritid fruit fly identification was undertaken in this study and the *Wolbachia* host status of colony-reared fruit flies was explored as a potential tool for species differentiation. The ability to identify multiple species simultaneously and rapidly without the need for costly downstream analysis and sequencing was deemed a priority. Multiplex PCR offers the ability to amplify different DNA targets and different amplicon sizes in a single run. Although the use of multiplex PCR for fruit fly identification has not been well explored, it has shown promising results in differentiating a species of interest, *Rhagoletis cerasi*, from other tephritid flies in North America as well as fruit fly parasitoid identification (Shariff *et al.*, 2014; Barr *et al.*, 2021). While the five fruit flies under study can be identified through a variety of existing molecular assays, to date, no assay can identify all five flies simultaneously. Therefore, this study utilises a multiplex PCR approach to provide a fast and accurate identification assay for differentiation of five tephritid fruit flies of economic importance to South Africa without life-stage restrictions. Many of these tephritid fruit fly species also occur in other parts of Africa (De Meyer, 2001; Virgilio *et al.*, 2013; De Meyer *et al.*, 2016; De Villiers *et al.*, 2016). As such, the development of a rapid and accurate identification technique in this study will be applicable for fruit fly identification in other parts of Africa where these species occur and are of economic importance.

4.2. Methods and materials

4.2.1. Sample collection, identification, and DNA extraction

Specimens used in this study were stored in 100% ethanol and kept at 4°C until used. Colony insects and larvae came from established colonies held at Citrus Research International (CRI) in Mbombela, Mpumalanga, South Africa. Detailed information regarding the origin of the colonies is listed in Supplementary Table S1. The identities of the fruit fly species in the colonies (adult specimens from colonies refreshed in the period 2020-2021) were confirmed by Marc De Meyer, Royal Museum for Central Africa, on 21 February 2022. DNA was extracted from single insects following an adapted "salting out" protocol by Sunnucks and Hales (1996), with TNES buffer (50 mM Tris, pH 7.5, 400 mM NaCl, 20 mM EDTA, 0.5 % SDS) substituted for 180 µl ATL buffer (Qiagen) and incubation taking place overnight at 56°C. Following the NaCl precipitation, 2 µl RNAse A was added to the supernatant and the second precipitation took place overnight at -20°C with isopropanol. DNA concentration and quality were quantified using a NanoDrop 2000 spectrophotometer and a Qubit dsDNA BR assay kit (Invitrogen).

Wild insects used for validation of the assay were collected from traps. Flies of the genus *Ceratitis* were trapped with McPhail type bucket traps baited with enriched ginger root oil (EGO lure) (Insect Science, Tzaneen, South Africa), and *B. dorsalis* flies were trapped with Chempac bucket traps baited with methyl eugenol (ME) (Invader lure, RiverBioscience, Gqeberha, South Africa). Total DNA was extracted from the whole body of the fruit fly following the destructive protocol of the DNeasy Blood and Tissue Kit (Qiagen).

The species of each adult colony specimen in this study was confirmed before the assay design using universal primer set CI-J2183 and TL2-N3014 (Simon *et al.*, 1994) for amplification and Sanger sequencing of the COI gene. The PCR was performed in a total volume of 25 µl containing 1x Kapa Taq buffer A (KAPA Biosystems), 0.2 mM dNTP mix (Thermo Scientific), 0.4 µM of each primer (CI-J2183 and TL2-N3014), and 0.05 U KAPA Taq DNA Polymerase (KAPA Biosystems). The cycling conditions included an initial denaturation step at 94°C for 5 min, followed by 35 cycles of 94°C for 30 s, annealing at 50°C for 30 s and elongation at 72°C for 45 s. The final extension took place at 72°C for 7 min.

4.2.2. High Throughput Sequencing and De novo assembly

DNA from two adult male specimens from the colony of each species were sent for high throughput sequencing at Macrogen (South Korea). Macrogen conducted library construction and high throughput sequencing of the colony insects on the Illumina NovaSeq 6000 platform (2 x 150 bp paired-end reads). Library preparation was performed using the TruSeq DNA PCR-Free Kit for the samples *C. rosa* 1, *C. quilicii* 1 & 2 and *C. cosyra* 2; and the TruSeq Nano DNA Kit for samples *C. capitata* 1 & 2, *C. rosa* 1, *C. cosyra* 1, and *B. dorsalis* 1 & 2, with input ranging from 0.565 to 2.998 µg of genomic DNA. De novo assembly was performed using CLC genomics workbench version 11.0.1 (Qiagen) and SPAdes (Nurk *et al.*, 2013) using default parameters as well as Velvet (Zerbino, 2010) with a hash length of 55.

4.2.3. Gene selection

Gene regions frequently used for differentiation of insect species were selected from literature and underwent preliminary bioinformatic analyses. A detailed list of these genes is available in Supplementary Table S2. The de novo assembled contigs were queried using BLAST+ standalone (BLASTn algorithm) against a local copy of the NCBI GenBank nucleotide database. The gene regions of interest were then identified, and multiple sequence alignments were constructed to compare the genes between species using CLC genomics workbench version 11.0.1 (Qiagen). De novo assembled contigs with high similarity to GenBank accessions XM_004526176.3 and XM_011215866.3 (annotated as Opsin Rh3/Rh4) were targeted for species differentiation and primer design. This gene region showed the greatest potential for species identification due to the number of single nucleotide polymorphisms observed between species in the multiple sequence alignment. Literature suggests that the function of opsins within the order Diptera extends beyond visual processes influencing adaptation to new ecological niches and playing additional roles in host fruit detection, gustatory reception, and taste (Feuda *et al.*, 2016, 2021; Papanicolaou *et al.*, 2016; Leung *et al.*, 2020; Sondhi *et al.*, 2021).

4.2.4. Primer design and multiplex PCR

A multiple sequence alignment of two reference sequences available on GenBank belonging to *C. capitata* and *B. dorsalis*, accessions XM_004526176.3 and XM_011215866.3, respectively as well as de novo-assembled contigs high in similarity to these reference sequences for each species (GenBank accessions: ON505377 – ON505386), was constructed. Five primer sets (IDT) were designed for differentiation of each species by amplicon size using Oligo Explorer 1.1.2 (Gene Link) (Table 4.1).

Table 4.1. List of primers designed for accurate species identification in the multiplex PCR assay.

Primer Pair	Sequence (5'-3')	Amplicon size (bp)
Opsin4_capitata_F	GCTAAAGCCATAACAATTCAG	327
Opsin4_capitata_R	CAGACTGTTCTTTTGGGC	
Opsin4_cosyra_F	GCTGTGACTTTGTTACAG	183
Opsin4_cosyra_R	GCATACTTGAATCTCAATCGAA	
Opsin4_quilicii_F	GCGTTCTGTTTTTAATCACTCA	128
Opsin4_quilicii_R	CATTTAATGTTTCAGAAGTGCT	
Opsin4_rosa_F	ATTGCTACAACTTTGTCGC	249
Opsin4_rosa_R	GCAGTAATACTGCGAATCATC	
Opsin4_dorsalis_F	TAGCACAATTATTTAGCGGG	676
Opsin4_dorsalis_R	ATTACCGTCAGCGATCAG	

The PCR was performed in a total volume of 25 µl containing 1x KAPA Taq buffer A (KAPA Biosystems), 0.2 mM dNTP mix (Thermo Scientific), 0.24 µM Opsin4_capitata_F & R, 0.32 µM Opsin4_cosyra_F & R, 0.32 µM Opsin4_quilicii_F & R, 0.64 µM Opsin4_rosa_F & R, 0.64 µM Opsin4_dorsalis_F & R and 0.05 U KAPA Taq DNA Polymerase (KAPA Biosystems). The cycling conditions included an initial denaturation step at 94°C for 5 min, followed by 35 cycles of 94°C for 30 s, annealing at 55°C for 30 s and elongation at 72°C for 35 s. The final extension took place at 72°C for 7 min.

All visualisations of multiplex PCR amplicons in this study were separated on a 2% agarose TAE (2 M Tris, 1 M glacial acetic acid, 0.05 M Na_2EDTA, pH 8) gel stained with ethidium bromide.

To confirm that each specific primer pair generated the expected amplicons, each amplicon was bi-directionally Sanger sequenced with the relevant species-specific primer pair at the Central Analytical Facility of Stellenbosch University. A dilution series with a dilution factor of 5 was made with DNA extracted from both adult and larval colony specimens to determine the assay's sensitivity. The dilution series was quantified using the Qubit dsDNA BR assay kit (Invitrogen). The multiplex PCR assay was thereafter performed with the dilution series for adult insect DNA (20 ng – 0.0064 ng) and for larval DNA (50 ng – 0.0000256 ng) to determine the limit of detection.

4.2.5. Assay Validation

The assay was validated using freshly extracted DNA from wild, trap-collected specimens of all five fruit flies, morphologically identified to species level using taxonomic keys. In total, the assay was validated on 15 wild fruit flies of each species. 1 µl DNA was taken directly from the extract and used in the multiplex PCR without normalisation for DNA concentration.

4.2.6. Determining *Wolbachia* host infection

De novo assembled contigs generated from HTS data were subjected to a local BLASTn analysis, where contigs were queried against a database containing reference sequences for multiple *Wolbachia* strains (Table S3). HTS sequencing reads were mapped to a reference genome of *Wolbachia pipientis* (CP041924.1) with the Burrows-Wheeler Aligner (BWA version) 7.0.13 (Li and Durbin, 2009). The resulting genome coverage of the mapping was evaluated using samtools (Danecek et al., 2021).

4.3. Results

4.3.1. Species identification and DNA extraction

All adult specimens used in this study underwent morphological identification and DNA barcoding targeting the COI gene with the universal primer pair CI-J2183 and TL2-N3014 (Simon *et al.*, 1994). All adult specimens were identified to species level through morphological identification using published keys (Virgilio, White and De Meyer, 2014). When DNA barcoding was carried out on these specimens, the COI region could only identify *C. capitata*, *C. cosyra*, and *B. dorsalis* to species level. Sequence similarity between *C. rosa* and *C.*

quilicii prevented differentiation based upon this gene region. DNA was successfully extracted from each specimen. DNA concentrations ranged from 25.4 - 320.0 ng/µl as determined by a Qubit dsDNA BR assay kit (Invitrogen). The DNA quality determined at the A260/A280 absorbance ratio on a NanoDrop 2000 spectrophotometer ranged between 1.9-2.12.

4.3.2. HTS assembly and BLASTn results

All three assembly platforms used, SPAdes, Velvet, and CLC, produced comparable assembly statistics for all ten specimen datasets evaluated (Table 4.2.). The N50 values of the assemblies were relatively low, ranging between 1263 bp – 8206 bp, which is expected when using default assembly parameters and not discarding smaller contigs (< 1000 bp). However, low N50 values have been reported previously in whole genome fruit fly assemblies, as was the case with the Medfly (Papanicolaou *et al.*, 2016) with an initial N50 of 3500 bp. The size of the largest contigs generated are comparable with those described in the whole genome sequencing of *Drosophila* spp. (Bronski *et al.*, 2020). BLASTn analysis revealed that the majority of the assembled contigs are of the host genus sequenced (either *Ceratitis* or *Bactrocera*). Of these contigs, the mitogenomic contigs were far less abundant than the genomic contigs (Table 4.3.). Very few viral and bacterial contigs were recovered from the dataset, with *B. dorsalis* (1 & 2) harbouring the greatest number of bacterial contigs overall across assembly platforms.

Table 4.2. Summary of assembly statistics for de novo assembly of HTS data of ten fruit fly specimens on three different assembly platforms.

Assembly platform	Specimen	No. contigs assembled	N50 (bp)	Largest contig (bp)
SPAdes	*C. capitata (1)*	1036666	8206	383489
	C. capitata (2)	1039438	7672	356350
	C. cosyra (1)	834903	4689	814363
	C. cosyra (2)	1036029	3758	814513
	C. quilicii (1)	304640	4381	1159991
	C. quilicii (2)	364764	3539	1378264
	B. dorsalis (1)	1550319	4784	801927
	B. dorsalis (2)	1818951	4490	1060960
Velvet	*C. rosa (1)*	2240849	1843	318799
	C. rosa (2)	2224798	1804	639153
CLC	*C. capitata (1)*	197175	6926	430393
	C. capitata (2)	215001	6710	430370
	C. cosyra (1)	245793	5265	646518
	C. cosyra (2)	286537	4470	834480
	C. quilicii (1)	479544	1688	1315181
	C. quilicii (2)	461375	1263	2324575
	C. rosa (1)	264686	5355	729438
	C. rosa (2)	282497	4914	594563
	B. dorsalis (1)	261949	5202	2546645
	B. dorsalis (2)	279559	5018	1051313

Table 4.3. Summary of BLASTn results for de novo assembled HTS contigs of ten fruit fly specimens on three different assembly platforms.

Assembly platform	Specimen	No. contigs assembled	No. host contigs	No. mitochondrial contigs	No. genomic contigs	No. viral contigs	No. viral contigs (>100 bp)	No. bacterial contigs	No. bacterial contigs (> 100 bp)
SPAdes	*C. capitata (1)*	262562	190286	2168	188118	57	13	52	40
	C. capitata (2)	281144	203934	2123	201811	41	9	2182	2176
	C. cosyra (1)	236523	154355	947	153408	171	111	1176	1170
	C. cosyra (2)	254159	161398	1081	160317	172	105	69	44
	C. quilicii (1)	128298	78054	346	77708	806	357	8	2
	C. quilicii (2)	130834	81051	401	80650	668	279	25	15
	B. dorsalis (1)	625599	539301	3113	536188	819	357	2767	2758
	B. dorsalis (2)	684148	584633	3528	581105	987	482	1031	1022
Velvet	*C. rosa (1)*	496640	365436	5631	359805	776	364	28	53
	C. rosa (2)	489697	347896	5619	342277	822	333	34	51
CLC	*C. capitata (1)*	82293	51617	387	51230	38	10	37	30
	C. capitata (2)	91635	53194	385	52809	38	9	834	834
	C. cosyra (1)	99589	56354	540	55814	90	52	268	260
	C. cosyra (2)	118477	61390	568	60822	90	45	63	45
	C. quilicii (1)	115651	68230	931	67299	469	260	14	3
	C. quilicii (2)	99205	58267	552	57715	353	205	20	11
	C. rosa (1)	117952	78067	680	77387	175	111	5	5
	C. rosa (2)	130447	81984	742	81242	169	87	25	21
	B. dorsalis (1)	128531	94171	725	93446	233	174	1809	1799
	B. dorsalis (2)	135453	98311	735	97576	207	144	1332	1330

4.3.3. Gene selection, primer design, specificity, and sensitivity

De novo assembled contigs with high similarity to GenBank accessions XM_004526176.3 and XM_011215866.3 were targeted for species differentiation and primer design. One primer set was designed for each species with differing amplicon lengths for use in a multiplex PCR. Specificity tests performed on freshly extracted DNA from colony-reared insects showed the presence of a single amplicon of the expected size for each species (Fig. 4.1). The results were consistent when tested on colony larvae, as shown in Fig. 4.2. A 2% agarose-TAE gel allowed for adequate separation of amplicons that were close in size for accurate species identification. Overall, the assay's detection limit was 10 ng and 4 ng when tested on colony larvae and colony adult DNA, respectively. Gel images of the sensitivity assay are presented in Supplementary Figure S2 and S3. At first, the multiplex PCR assay was designed to include an internal control amplicon (COI – amplified with the primer pair CI-J2183 and TL2-N3014), however, this internal control amplicon interfered with the reliability of the *C. cosyra*-specific amplicon and was thus excluded from the final multiplex protocol. Images of the assay with the internal control amplicon are presented in Supplementary Figure S4.

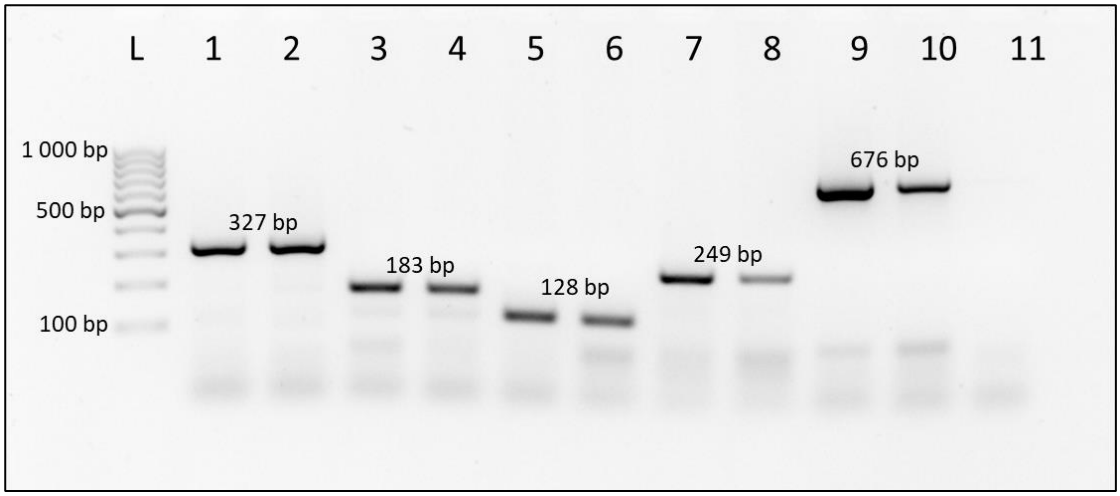


Fig. 4.1. 2% agarose-TAE gel displaying the specificity of the multiplex PCR assay on freshly extracted DNA from colony-reared insects with species-specific amplicon size indicated. Lanes 1 & 2: *C. capitata*, Lanes 3 & 4: *C. cosyra*, Lanes 5 & 6: *C. quilicii*, Lanes 7 & 8: *C. rosa*, Lanes 9 & 10: *B. dorsalis*, Lane 11: No template control, Lane L: GeneRuler 100 bp DNA ladder (Thermo Scientific). The original gel is presented in Supplementary Figure S1 (a.).

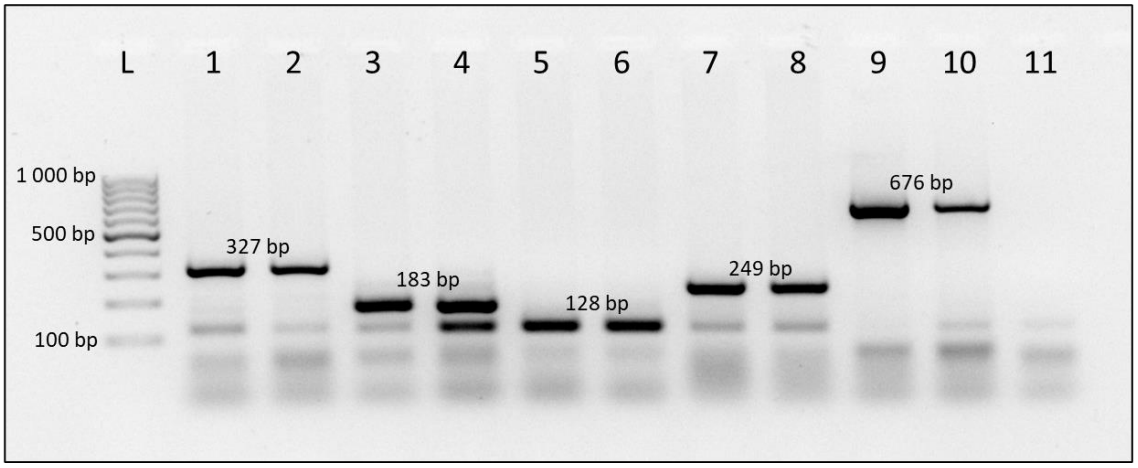


Fig. 4.2. 2% Agarose-TAE gel displaying the specificity of multiplex primers in the case of duplex formation in freshly extracted colony-reared larval DNA with species-specific amplicon size indicated. Lane 4 demonstrates the expected *C. cosyra* amplicon at 183 bp with non-specific amplification at 128 bp leading to the formation of a duplex; the larger 183 bp amplicon should be used for identification. Lane 5 is a single 128 bp amplicon indicative of *C. quilicii*. Lanes 1 & 2: *C. capitata*, Lanes 3 & 4: *C. cosyra*, Lanes 5 & 6: *C. quilicii*, Lanes 7 & 8: *C. rosa*, Lanes 9 & 10: *B. dorsalis*, Lane 11: No template control, Lane L: GeneRuler 100 bp DNA ladder (Thermo Scientific). The original gel is presented in Supplementary Figure S5 (a.).

4.3.4. Assay validation on wild insects

The multiplex PCR assay was validated using freshly extracted DNA from wild, trap-collected specimens morphologically identified using available taxonomic keys (Virgilio, White and De Meyer, 2014). It was noted that certain trap-collected specimens produced non-specific amplification of various sizes in addition to the expected identity amplicon. However, none of the non-specific amplicons interfered with the reliability or accuracy of the assay. An example of the efficacy of the multiplex PCR assay on trap-collected fruit flies can be found in Fig. 4.3.

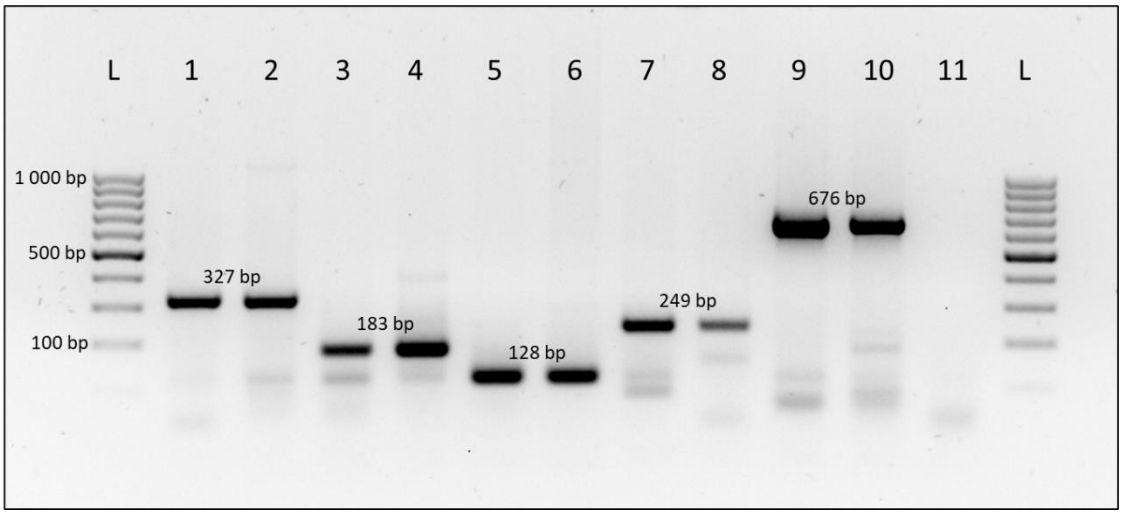


Fig. 4.3. 2% agarose-TAE gel displaying the efficacy of the Multiplex PCR assay to identify wild, trap-collected specimens with examples of non-specific amplification. Species-specific amplicon sizes are indicated. Lanes 1 & 2: *C. capitata*, Lanes 3 & 4: *C. cosyra*, Lanes 5 & 6: *C. quilicii*, Lanes 7 & 8: *C. rosa*, Lanes 9 & 10: *B. dorsalis*, Lane 11: No template control, Lane L: GeneRuler 100 bp DNA ladder (Thermo Scientific). The original gel is presented in Supplementary Figure S6.

4.3.5. *Wolbachia* host status

BLASTn results recovered small contigs, between 28 – 1745 bp, considering the genome size of *Wolbachia* (1 239 814 bp). Coverage statistics generated from the reference mapping of HTS reads for each fruit fly species against a reference genome for *Wolbachia pipientis* (CP041924.1) revealed that there is likely little to no bacterial contamination in these colony specimens, due to very low coverage and shallow coverage depth (Table 4.4). A minimum coverage breadth of 90% of the genome sequence is suggested for inferring a positive infection status of specimen (Richardson *et al.*, 2012; Signor, 2017).

Table 4.4. Coverage depth and breadth calculations for reference mapping of HTS reads against a reference genome of *Wolbachia pipientis* (CP041924.1) to determine host status.

Sample	No. of aligned reads	No. covered bases with depth >1	Coverage (% of covered bases)	Mean depth of coverage
C. capitata (1)	15169	6951	0.47	0.90
C. capitata (2)	27988	7224	0.49	1.71
C. cosyra (1)	66246	57539	3.88	3.25
C. cosyra (2)	57547	32553	2.19	3.32
C. quilicii (1)	36389	75695	5.10	19.73
C. quilicii (2)	783170	63847	4.30	45.91
C. rosa (1)	119294	89502	6.03	4.26
C. rosa (2)	135753	158670	10.69	3.86
B. dorsalis (1)	96046	98398	6.63	4.79
B. dorsalis (2)	60247	106128	7.15	2.33

4.2. Discussion

Five pairs of species-specific primers were designed, and a multiplex PCR was developed to identify fruit flies of economic importance in South Africa to species level. This assay generates a single amplicon of varying sizes for the different fruit fly species, *C. capitata* (327 bp), *C. cosyra* (183 bp), *C. quilicii* (128 bp), *C. rosa* (249 bp), and *B. dorsalis* (676 bp). These amplicons can be separated on a 2% agarose gel allowing for accurate differentiation without downstream analysis and sequencing. DNA concentrations of wild, trap-collected query specimens were not normalised during assay validation to demonstrate the efficacy of this assay for routine identification where concentration normalisation is not a priority, saving time when large numbers of specimens are being processed simultaneously.

In this study, all morphologically identified query specimens were correctly identified to species level using the multiplex PCR assay. This assay was developed for use as a differentiation tool to identify fruit fly pests of fresh fruit in South Africa that could potentially be present in export consignments and only validated to identify the five fruit fly pest species currently present in the country. The false-positive rate incurred when other fruit fly species are queried using this assay is unknown. Since other tephritid flies present in South Africa are not pests of commercial fruit primarily exported from South Africa, it is expected that only the five fly species investigated are likely to be intercepted on commercial fruit produced for export.

It was noted that in some specimens, a non-specific amplicon was present at the same size as the expected *C. quilicii* amplicon. The presence of this duplex is not to be confused with the *C. quilicii* amplicon, which will always yield a single 128 bp amplicon with no non-specific amplification. In cases where the duplex is present, the larger amplicon is to be used for species identification (Fig. 2.). The species-specific primer pair designed for *C. quilicii* is located just outside of the opsin Rh4 coding domain in an intergenic region which has the potential to cross amplify in closely related species. Although the location of this primer set can lead to false-positive amplification in closely related species when used in conjunction with the other four species-specific primer sets designed in this study in a multiplex, accurate species differentiation of the five flies investigated can be achieved. The universal primer set CI-J2183 and TL2-N3014 (Simon *et al.*, 1994) can be used to amplify the COI gene, which can be sequenced and queried against the NCBI database for identity confirmation where applicable. It should be noted that CI-J2183 and TL2-N3014 can accurately identify *C. capitata*, *C. cosyra,* and *B. dorsalis* to species level. *Ceratitis capitata* and *Ceratitis caetrata* Munro (Diptera: Tephritidae) have previously been shown to share a high sequence similarity within the COI gene region, which may result in erroneous identification (Frey *et al.*, 2013). However, the distribution of *C. caetrata* is limited to Kenya, and the fly has never been reported in South Africa (De Meyer, 2000). In closely related species and species complexes such as *C. rosa* and *C. quilicii,* this gene region does not allow accurate differentiation (De Meyer *et al.*, 2016; Dhami *et al.*, 2016; Zhang *et al.*, 2021).

While the multiplex PCR assay designed in this study reliably performs its role, the downside of a multiplex is related to the use of PCR itself, where the ability of the assay to reliably detect species present relies on the primer binding region to be conserved enough within the species so that any intraspecific variation present in the target region does not hinder amplification leading to false-negative results. This assay was validated with 15 wild insects per species collected from various sites across South Africa (Table S4), and no false-negative results were obtained. False-positive results can also occur where closely related species are highly similar to the target species leading to amplification. Fortunately, false-positive results are avoided in the multiplex assay described as species-specific amplicons are also size specific, so in cases of cross-amplification observed (Fig. 2.) where a duplex is formed, the larger amplicon is followed for accurate species identification. There is a high potential for false-positive results when other tephritid flies are queried against this assay, however, this assay is intended for use as a differentiation tool for identifying only the five fruit flies investigated.

The multiplex PCR detection assay developed in this study has application in identifying and monitoring agricultural pests of phytosanitary significance, both for pest management and surveillance practices. This relatively low-cost and easy-to-perform assay uses only essential molecular laboratory equipment. It can be used in a standalone format or in conjunction with existing morphological identification techniques for improved accuracy in species identification. A significant advantage of this proposed method is that it allows for identification to species level without the need for downstream analysis. Reliable species identification can be achieved in under two and a half hours post DNA extraction, which significantly reduces the time required for existing molecular identification by DNA barcoding (Armstrong and Ball, 2005). The increased turnaround time is a considerable advantage for inspection purposes in the implementation of a systems approach reducing the risk of fruit flies before fruit export, for inspections of fruit consignments at ports of entry as well as for early detection of invasive fruit flies such as *B. dorsalis* which is currently absent in several areas in South Africa (*Government Gazette*, 2017).

The HTS datasets generated during this study are rich sources of genetic information expanding beyond the genomes sequenced. As demonstrated above (Table 4.3.), the de novo assembled datasets contained both viral and bacterial sequence information as well. While the viral and bacterial contigs present were low relative to the size of the host genome, datasets such as these, can be mined and optimised in such a way to maximise the number of viral and bacterial and other microbial data present. This study focused on species molecular differentiation hence, analyses performed were not ideal for identifying species present in the database other than the host. Although no detectable *Wolbachia* infections were identified, *Wolbachia* to host DNA ratio is typically very low, so much so that previous studies which recovered complete genomes of *Wolbachia* from insect specimen have required the pooling of DNA from multiple host specimen (Richardson *et al.*, 2012; Signor, 2017). Low levels of *Wolbachia* to host DNA combined with the data analyses being optimised for species differentiation rather than microbe detection are potential causes for the lack of

Wolbachia infections identified. The use of existing *Wolbachia* specific primers (Braig *et al.*, 1998; Zhou, Rousset and O'Neill, 1998), on both colony-reared and wild-trap collected specimen, would be an interesting comparison and avenue to further the research in this field.

This study focused on fruit fly identification in the South African context to facilitate monitoring and inspection processes relating to fruit flies of economic importance. While species-specific *Wolbachia* strains were not detected, further research into endosymbionts as potential drivers of speciation and species-specific identification markers would be a valuable resource. A multiplex PCR was developed for accurate species identification of the five fruit flies of focus. Given that these species also occur in other parts of Africa and are of economic importance in these regions, the assay may be of practical use in these regions as well. Further research will be required to determine the suitability of this assay for fruit fly identification in other African countries where other economically important tephritid flies occur, more specifically other members of the *Ceratitis* FARQ complex (*Ceratitis fasciventris* and *Ceratitis anonae*). Presently, the datasets generated during this study contain useful information that can be optimised for greater use in microbe detection and other species-specific markers, and the multiplex PCR assay developed in this study will provide a useful aid in decision-making regarding international trade and for monitoring and detection purposes.

4.3. References

Armstrong, K. F. and Ball, S. L. (2005) 'DNA barcodes for biosecurity: invasive species identification', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1462), pp. 1813–1823. doi: 10.1098/rstb.2005.1713.

Balmès, V. and Mouttet, R. (2017) 'Development and validation of a simplified morphological identification key for larvae of tephritid species most commonly intercepted at import in Europe', *EPPO Bulletin*, 47(1), pp. 91–99. doi: 10.1111/epp.12369.

Barr, N. *et al.* (2006) 'Molecular diagnostics of economically important *Ceratitis* fruit fly species (Diptera: Tephritidae) in Africa using PCR and RFLP analyses', *Bulletin of Entomological Research*, 96, pp. 505–521. doi: 10.1079/BER2006452.

Barr, N. B. *et al.* (2012) 'Molecular Identification of *Ceratitis capitata* (Diptera: Tephritidae) Using DNA Sequences of the COI Barcode Region', *Annals of the Entomological Society of America*, 105(2), pp. 339–350. doi: 10.1603/AN11100.

Barr, N. B. *et al.* (2021) 'Using the rDNA Internal Transcribed Spacer 1 to Identify the Invasive Pest *Rhagoletis cerasi* (Diptera: Tephritidae) in North America', *Journal of Economic Entomology*. Edited by K. Shelby, 114(1), pp. 360–370. doi: 10.1093/jee/toaa281.

Boykin, L. M. *et al.* (2012) 'Species Delimitation and Global Biosecurity', *Evolutionary Bioinformatics*, 8, p. EBO.S8532. doi: 10.4137/EBO.S8532.

Bragard, C. *et al.* (2020) 'Pest categorisation of non-EU Tephritidae', *EFSA Journal*, 18(1). doi: 10.2903/j.efsa.2020.5931.

Braig, H. R. *et al.* (1998) 'Cloning and Characterization of a Gene Encoding the Major Surface Protein of the Bacterial Endosymbiont *Wolbachia pipientis*', *Journal of Bacteriology*, 180(9), pp. 2373–2378. doi: 10.1128/JB.180.9.2373-2378.1998.

Bronski, M. J. *et al.* (2020) 'Whole Genome Sequences of 23 Species from the *Drosophila montium* Species Group (Diptera: Drosophilidae): A Resource for Testing Evolutionary Hypotheses', *G3 Genes|Genomes|Genetics*, 10(5), pp. 1443–1455. doi: 10.1534/g3.119.400959.

Delatte, H. *et al.* (2013) 'Isolation and characterisation of sixteen microsatellite markers cross-amplifying in a complex of three African agricultural pests (*Ceratitis rosa*, *C. anonae* and *C. fasciventris*, Diptera: Tephritidae)', *Conservation Genetics Resources*, 5(1), pp. 31–34. doi: 10.1007/s12686-012-9722-6.

Delatte, H. *et al.* (2014) 'Isolation and characterisation of sixteen microsatellite markers amplifying an African agricultural pest, *Ceratitis cosyra* (Walker) (Diptera: Tephritidae)', *Conservation Genetics Resources*, 6(1), pp. 9–11. doi: 10.1007/s12686-013-0026-2.

De Meyer, M. (2000) 'Systematic revision of the subgenus *Ceratitis* MacLeay s.s. (diptera, Tephritidae)', *Zoological Journal of the Linnean Society*, 128, pp. 439–467. doi: 10.1006/zils. 1999.01 98.

De Meyer, M. (2001) 'Distribution patterns and host-plant relationships within the genus *Ceratitis* MacLeay (Diptera: Tephritidae) in Africa', *Cimbebasia*, 17, pp. 219–228.

De Meyer, M. *et al.* (2016) 'Description of new *Ceratitis* species (Diptera: Tephritidae) from Africa, or how morphological and DNA data are complementary in discovering unknown species and matching sexes', *European Journal of Taxonomy*, 233(233), pp. 1–23. doi: 10.5852/ejt.2016.233.

De Villiers, M. *et al.* (2013) 'The Distribution, Relative Abundance, and Seasonal Phenology of *Ceratitis capitata* , *Ceratitis rosa* , and *Ceratitis cosyra* (Diptera: Tephritidae) in South Africa', *Environmental Entomology*, 42(5), pp. 831–840. doi: 10.1603/EN12289.

De Villiers, M. *et al.* (2016) 'The potential distribution of *Bactrocera dorsalis* : considering phenology and irrigation patterns', *Bulletin of Entomological Research*, 106(1), pp. 19–33. doi: 10.1017/S0007485315000693.

Dhami, M. K. *et al.* (2016) 'A real-time PCR toolbox for accurate identification of invasive fruit fly species', *Journal of Applied Entomology*, 140(7), pp. 536–552. doi: 10.1111/jen.12286.

EPPO (2020) 'PM 3/90 (1) Inspection of citrus fruits consignments', *EPPO Bulletin*, 50(3), pp. 383–400. doi: 10.1111/epp.12684.

Feuda, R. *et al.* (2016) 'Conservation, Duplication, and Divergence of Five Opsin Genes in Insect Evolution', *Genome Biology and Evolution*, 8(3), pp. 579–587. doi: 10.1093/gbe/evw015.

Feuda, R. *et al.* (2021) 'Phylogenomics of Opsin Genes in Diptera Reveals Lineage-Specific Events and Contrasting Evolutionary Dynamics in *Anopheles* and *Drosophila*', *Genome Biology and Evolution*. Edited by A. Betancourt, 13(8). doi: 10.1093/gbe/evab170.

Frey, J. E. *et al.* (2013) 'Developing diagnostic SNP panels for the identification of true fruit flies (Diptera: Tephritidae) within the limits of COI-based species delimitation', *BMC Evolutionary Biology*, 13(1), p. 106. doi: 10.1186/1471-2148-13-106.

Government Gazette (2017) 'Agricultural Pests Act, 1983 (Act No. 36 of 1983) Control measures: Amendment, in: Department of Agriculture, Forestry and Fisheries', pp. 11–17.

Van Houdt, J. K. J. *et al.* (2010) 'Recovering full DNA barcodes from natural history collections of Tephritid fruitflies (Tephritidae, Diptera) using mini barcodes', *Molecular Ecology Resources*, 10(3), pp. 459–465. doi: 10.1111/j.1755-0998.2009.02800.x.

Jiang, F. *et al.* (2014) 'Existence of species complex largely reduced barcoding success for invasive species of Tephritidae: a case study in *Bactrocera* spp.', *Molecular Ecology Resources*, 14(6), pp. 1114–1128. doi: 10.1111/1755-0998.12259.

Jiang, F. *et al.* (2016) 'A high-throughput detection method for invasive fruit fly (Diptera: Tephritidae) species based on microfluidic dynamic array', *Molecular Ecology Resources*, 16(6), pp. 1378–1388. doi: 10.1111/1755-0998.12542.

Jiang, F. *et al.* (2018) 'A conserved motif within cox 2 allows broad detection of economically important fruit flies (Diptera: Tephritidae)', *Scientific Reports*, 8(1), p. 2077. doi: 10.1038/s41598-018-20555-2.

Leung, N. Y. *et al.* (2020) 'Functions of Opsins in *Drosophila* Taste', *Current Biology*, 30(8), pp. 1367-1379.e6. doi: 10.1016/j.cub.2020.01.068.

Louzeiro, L. R. F. *et al.* (2021) 'Incidence of frugivorous flies (Tephritidae and Lonchaeidae), fruit losses and the dispersal of flies through the transportation of fresh fruit', *Journal of Asia-Pacific Entomology*, 24(1), pp. 50–60. doi: 10.1016/j.aspen.2020.11.006.

Malacrida, A. R. *et al.* (2007) 'Globalization and fruitfly invasion and expansion: the medfly paradigm', *Genetica*, 131(1), pp. 1–9. doi: 10.1007/s10709-006-9117-2.

Manrakhan, A., Venter, J. H. and Hattingh, V. (2015) 'The progressive invasion of *Bactrocera dorsalis* (Diptera: Tephritidae) in South Africa', *Biological Invasions*, 17(10), pp. 2803–2809. doi: 10.1007/s10530-015-0923-2.

Nurk, S. *et al.* (2013) 'Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads', in Deng, M. et al. (eds) *Research in Computational Molecular Biology*. Berlin: Springer, Berlin, Heidelberg, pp. 158–170. doi: 10.1007/978-3-642-37195-0_13.

Papanicolaou, A. *et al.* (2016) 'The whole genome sequence of the Mediterranean fruit fly, *Ceratitis capitata* (Wiedemann), reveals insights into the biology and adaptive evolution of a highly invasive pest species', *Genome Biology*, 17(1), p. 192. doi: 10.1186/s13059-016-1049-2.

Prezotto, L. F. *et al.* (2017) '*Wolbachia* strains in cryptic species of the *Anastrepha fraterculus* complex (Diptera, Tephritidae) along the Neotropical Region', *Systematic and Applied Microbiology*, 40(1), pp. 59–67. doi: 10.1016/j.syapm.2016.11.002.

Qin, Y. *et al.* (2018) 'Population structure of a global agricultural invasive pest, *Bactrocera dorsalis* (Diptera: Tephritidae)', *Evolutionary Applications*, 11(10), pp. 1990–2003. doi: 10.1111/eva.12701.

Rasolofoarivao, H., Raveloson Ravaomanarivo, L. H. and Delatte, H. (2022) 'Host plant ranges of fruit flies (Diptera: Tephritidae) in Madagascar', *Bulletin of Entomological Research*, 112(1), pp. 1–12. doi: 10.1017/S0007485321000511.

Riegler, M. and Stauffer, C. (2008) '*Wolbachia* infections and superinfections in cytoplasmically incompatible populations of the European cherry fruit fly *Rhagoletis cerasi* (Diptera, Tephritidae)', *Molecular Ecology*, 11(11), pp. 2425–2434. doi: 10.1046/j.1365-294X.2002.01614.x.

Shariff, S. *et al.* (2014) 'Multiplex PCR in determination of Opiinae parasitoids of fruit flies, *Bactrocera* sp., infesting star fruit and guava', *Journal of Insect Science*. Edited by K. Iatrou, 14(1). doi: 10.1093/jis/14.1.7.

Simon, C. *et al.* (1994) 'Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers', *Annals of the Entomological Society of America*, 87(6), pp. 651–701. doi: 10.1093/aesa/87.6.651.

Sondhi, Y. *et al.* (2021) 'Light environment drives evolution of color vision genes in butterflies and moths', *Communications Biology*, 4(1), p. 177. doi: 10.1038/s42003-021-01688-z.

Sun, X., Cui, L. and Li, Z. (2007) 'Diversity and phylogeny of wolbachia infecting *Bactrocera dorsalis* (Diptera: Tephritidae) populations from China', *Environmental Entomology*, 36(5), pp. 1283–1289.

Sunnucks, P. and Hales, D. F. (1996) 'Numerous Transposed Sequences of Mitochondrial Cytochrome Oxidase I-II in Aphids of the Genus *Sitobion* (Hemiptera: Aphididae)', *Molecular biology and evolution*, 13(3), pp. 510–524. doi: 10.1093/oxfordjournals.molbev.a025612.

Tan, K. H. *et al.* (2013) 'Comparison of methyl eugenol metabolites, mitochondrial COI, and rDNA sequences of *Bactrocera philippinensis* (Diptera: Tephritidae) with those of three other major pest species within the *dorsalis* complex', *Applied Entomology and Zoology*, 48(3), pp. 275–282. doi: 10.1007/s13355-013-0183-5.

Virgilio, M. *et al.* (2008) 'Molecular evaluation of nominal species in the *Ceratitis fasciventris*, *C. anonae*, *C. rosa* complex (Diptera: Tephritidae)', *Molecular Phylogenetics and Evolution*, 48(1), pp. 270–280. doi: 10.1016/j.ympev.2008.04.018.

Virgilio, M. *et al.* (2012) 'Identifying Insects with Incomplete DNA Barcode Libraries, African Fruit Flies (Diptera: Tephritidae) as a Test Case', *PLoS ONE*. Edited by D. Steinke, 7(2), p. e31581. doi: 10.1371/journal.pone.0031581.

Virgilio, M. *et al.* (2013) 'Cryptic diversity and gene flow among three African agricultural pests: *Ceratitis rosa* , *Ceratitis fasciventris* and *Ceratitis anonae* (Diptera, Tephritidae)', *Molecular Ecology*, 22(9), pp. 2526–2539. doi: 10.1111/mec.12278.

Virgilio, M. *et al.* (2019) 'An integrated diagnostic setup for the morphological and molecular identification of the *Ceratitis* FAR complex (*C. anonae*, *C. fasciventris*, *C. rosa*, *C. quilicii*, Diptera, Tephritidae)', *Bulletin of Entomological Research*, 109(3), pp. 376–382. doi: 10.1017/S0007485318000615.

Virgilio, M., White, I. and De Meyer, M. (2014) 'A set of multi-entry identification keys to African frugivorous flies (Diptera, Tephritidae)', *ZooKeys*, 428, pp. 97–108. doi: 10.3897/zookeys.428.7366.

Weeks, A. R., Reynolds, T. K. and Hoffmann, A. A. (2002) '*Wolbachia* dynamics and host effects: what has (and has not) been demonstrated?', *Trends in Ecology & Evolution*, 17(6), pp. 257–262. doi: 10.1016/S0169-5347(02)02480-1.

Whatson, M. (2020) 'Decision To Revise Import Requirements for the Importation of Fresh Citrus From South Africa Into the United States', *Federal Registar*, 85(215).

White, I. M. and Elson-Harris, M. M. (1992) *Fruit flies of economic significance: their identification and bionomics*. CAB international.

Zerbino, D. R. (2010) 'Using the Velvet de novo Assembler for Short-Read Sequencing Technologies', *Current Protocols in Bioinformatics*, 31(1). doi: 10.1002/0471250953.bi1105s31.

Zhang, Y. *et al.* (2021) 'Phylogenomic resolution of the *Ceratitis* FARQ complex (Diptera: Tephritidae)', *Molecular Phylogenetics and Evolution*, 161, p. 107160. doi: 10.1016/j.ympev.2021.107160.

Zhou, W., Rousset, F. and O'Neill, S. (1998) 'Phylogeny and PCR–based classification of *Wolbachia* strains using wsp gene sequences', *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 265(1395), pp. 509–515. doi: 10.1098/rspb.1998.0324.

CHAPTER 5. CONCLUSION

5.2. Research summary

Fresh fruit production is an important export-focused industry in South Africa, representing over a third of the country's total agricultural exports. Future economic and agricultural development in this BRICS nation is dependent on the shift to more sustainable agriculture. Decreasing food wastage of fresh fruit is an attainable goal for the industry. Tephritid fruit flies cause both direct and indirect fruit loss from oviposition damage, blind stings, and quarantine restrictions on export markets (Badii *et al.*, 2015). *Ceratitis capitata*, *C. cosyra*, *C. rosa*, *C. quilicii*, and *B. dorsalis* are the five major fruit fly pests affecting the production and export of fresh fruit in South Africa. Although fruit fly management strategies and protocols are in place, effective and accurate fruit fly monitoring is still necessary for assessing the success of these strategies, detecting invasions, and making quarantine decisions about the fate of consignments at PoE. Current fruit fly identification is performed using morphology based taxonomic keys, and where uncertainties arise (as a result of larval interceptions, or damaged specimens), DNA barcoding is used. However, this molecular method cannot distinguish between closely related or cryptic species (Jiang *et al.*, 2014; De Meyer *et al.*, 2016). Although other molecular detection assays have been developed for fruit fly identification, none have been optimised for use in the South African context. Therefore, the purpose of this research study was to develop a molecular identification assay that can accurately differentiate between these five flies simultaneously and timeously.

In the first research chapter (chapter 3), ten complete mitochondrial genomes were constructed from HTS data. A primer set (Mito_F/R) was designed to amplify a DNA fragment between $tRNA^{ile}$ and $tRNA^{met}$. Within this amplicon, an intergenic region spanning $tRNA^{ile}$ – $tRNA^{gln}$ (designated intergenic region I) was found to be species-specific in size. The length differences of intergenic region I is proposed as a species differentiation method, where *C. capitata* (40 bp), *C. cosyra* (2 bp), *C. quilicii* (62-64 bp), *C. rosa* (34 bp) and *B. dorsalis* (0 bp) can be identified via a sequencing-based assay. Since this assay takes roughly four days for results and provides valuable sequencing data, it would be best suited for pest surveillance during routine monitoring. Sequencing facilities are not always readily available at PoE, and the time delay for identification results can be costly. Sequencing data provided from this assay can be used to survey for these fruit flies and infer evolutionary and phylogenetic phenomena.

This study also investigated the use of HRM analysis of mitochondrial intergenic regions for species differentiation. However, the low GC% of the selected region used in this study was unsuitable for HRM analysis as consistent dye intercalation was affected. Hence, melt-point confidence intervals for each species were inconsistent and could not be accurately determined.

In the second research chapter (chapter 4), a second identification assay was developed whereby HTS data was de novo assembled and the gene opsin Rh4 was targeted for species-specific primer design. Five sets of

species-specific primers were designed, and an end-point multiplex PCR assay was optimised for rapid identification of these fruit flies. This assay is able to differentiate between the five fruit fly species based on species-specific amplicon size: *C. capitata* (327 bp), *C. cosyra* (183 bp), *C. quilicii* (128 bp), *C. rosa* (249 bp), and *B. dorsalis* (676 bp), on a 2% agarose-TAE gel. This assay provides results in 2.5 hours (post DNA extraction). Given the multiplex assay's rapid turnaround time and use of basic laboratory equipment, it would be best suited for use in time-sensitive applications such as larval interceptions at PoE. The endosymbiont, *Wolbachia pipientis*, was also considered as a potential speciation mechanism whereby different fruit fly species potentially harbour species-specific *Wolbachia* strains (Keeling, Jiggins and Read, 2003). However, the *Wolbachia* host status of the fruit flies in this study was negative. This either indicates that the fruit fly colonies are free from bacterial contamination or that the endosymbiont-to-host DNA ratio was too low to detect.

5.3. Considerations for future research

While the current research study only focused on the identification and differentiation of five major fruit fly pests in South Africa, future research should aim to improve the assay by including non-major tephritid pests present in the country, major tephritid pests in neighbouring countries that pose an invasion risk such as *Drosophila suzukii*, as well as the common vinegar fly (*Drosophila* spp.) which, although pose no phytosanitary risk, are often mistaken for tephritid larvae in packhouses. Such improvements need not reinvent the wheel but can contribute additively to the multiplex PCR assay described (chapter 4) through designing species-specific primers for these additional flies.

The continuous generation of sequence data, both high throughput and Sanger, for non-model organisms such as Tephritidae is of utmost importance for a deeper understanding of this complex fruit fly family. Expanding existing databases improves the phylogenetic and evolutionary understanding of these species. Mining existing HTS datasets, such as those generated in this study, can reveal new insights into these species and define new molecular markers.

Endosymbionts present in the gut microbiome of these fruit flies have been explored for use in biological control applications and can potentially represent a mechanism of speciation (Boller *et al.*, 1976). Although the data generated in this study were negative for *Wolbachia* infections, further research into *Wolbachia* host dynamics would be valuable. Particularly regarding their potential role in speciation in the case of cryptic species *C. rosa* and *C. quilicii* and towards the application of the incompatible insect technique (IIT).

Sanger sequencing data generated with the primer set Mito_F/R (chapter 3) should be stored and used to monitor potential population or evolutionary changes in these fruit fly species. It would be valuable to track any abnormalities within *C. quilicii*, as populations of this newly defined species tend to be unstable. Further research should consider optimising the primer set Mito_F/R for capillary electrophoresis separation, where a one bp difference in amplicon length can be detected.

5.4. Concluding remarks

Research into molecular methods for agricultural pest identification and detection is rising along with the demand for accuracy and reliability in diagnostics. Continual trade growth and globalisation increase the risk of spread of invasive and potentially destructive species. This study has accomplished its aim and objectives in developing two identification assays applicable for the differentiation of five fruit flies of economic importance in South Africa. These assays will improve fruit fly monitoring practices and can accurately and timeously contribute to decisions of quarantine and phytosanitary significance at PoE. The impact of such molecular tools can indirectly lead to reductions in food wastage and directly improve market access for the export of South African fresh fruit through the advancement in economically important species detection.

5.5. References

Badii, K. B. *et al.* (2015) 'Review of the pest status, economic impact and management of fruit-infesting flies (Diptera: Tephritidae) in Africa', *African Journal of Agricultural Research*, 10(12), pp. 1488–1498. doi: 10.5897/AJAR2014.9278.

Boller, E. F. *et al.* (1976) 'Incompatible races of European cherry fruit fly, *Rhagoletis cerasi* (Diptera: Tephritidae), their origin and potential use in biological control', *Entomologia Experimentalis et Applicata*, 20(3), pp. 237–247. doi: 10.1111/j.1570-7458.1976.tb02640.x.

Jiang, F. *et al.* (2014) 'Existence of species complex largely reduced barcoding success for invasive species of Tephritidae: a case study in *Bactrocera* spp.', *Molecular Ecology Resources*, 14(6), pp. 1114–1128. doi: 10.1111/1755-0998.12259.

Keeling, M. J., Jiggins, F. M. and Read, J. M. (2003) 'The invasion and coexistence of competing *Wolbachia* strains', *Heredity*, pp. 382–388.

De Meyer, M. *et al.* (2016) 'Description of new *Ceratitis* species (Diptera: Tephritidae) from Africa, or how morphological and DNA data are complementary in discovering unknown species and matching sexes', *European Journal of Taxonomy*, (233). doi: 10.5852/ejt.2016.233.

APPENDIX A. SUPPLEMENTARY MATERIAL FOR CHAPTER 3

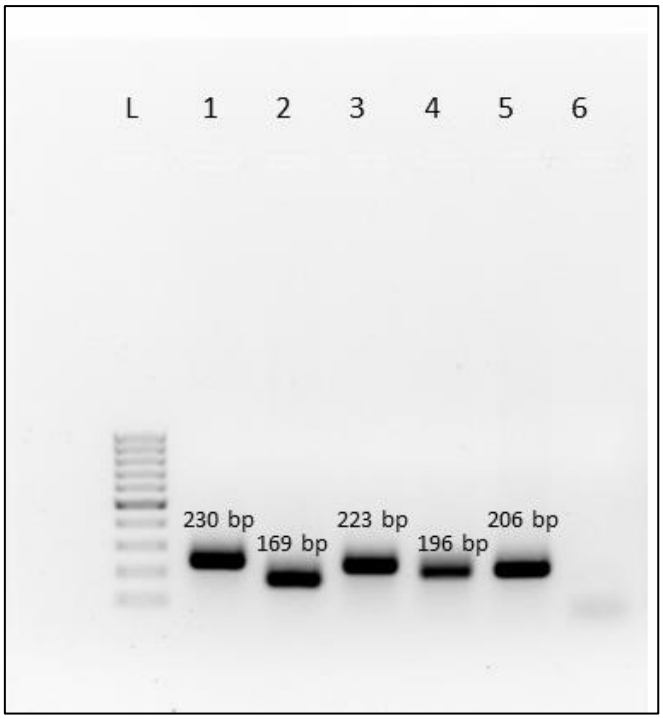


Figure S1. 2% agarose TAE gel visualised with ethidium bromide displaying specificity of the primer pair Mito_F/R. Lane 1: *C. capitata*, Lane 2: *C. cosyra*, Lane 3: C*. quilicii*, Lane 4: *C. rosa*, Lane 5: *B. dorsalis*, Lane 6: no template control, Lane L: 100 bp DNA ladder (Thermo Scientific). Amplicon sizes are indicated on the gel.

Table S1. Collection information of colony flies and respective larvae reared at CRI (Mbombela, Mpumalanga, South Africa). Initial collection sites of the established colonies are provided as coordinates. Adult colony insects were collected in February 2021; these colonies were refreshed between January 2019 and January 2020. Larval specimens were collected in August 2021; these colonies were refreshed between November 2020 and May 2021.

Life stage	Species	Rearing fruit		Collection Date	Latitude	Longitude
Adult male	*C. capitata*	Coffee	(*Coffea canephora Pierre ex. Froehner)*	10/03/2019	31°5’15.63” E	25°6’43.50” S
	C. cosyra	Marula	(*Sclerocarya birrea* (A. Rich.) Hochst.)	1/29/2019	31°2’35.70” E	25°28’4.03” S
	C. quilicii	Peach	(*Prunus persica* L. Batsch)	1/21/2020	30°23’34.73” E	24°59’47.17” S
	C. rosa	Strawberry Guava	(*Psidium cattleyanum* Sabine)	11/13/2019	30°58’10.99’’ E	25°27’08.54’’ S
	B. dorsalis	Mango	(*Mangifera indica* L.)	1/29/2019	30°58’10.99’’ E	25°27’08.54’’ S
Larvae	*C. capitata*	Coffee	(*Coffea canephora Pierre ex.* Froehner)	3/10/2021	31°5’15.63” E	25°6’43.50” S
	C. cosyra	Pepper-bark tree	(*Warburgia salutaris* (Bertol.f.) Chiov.)	11/30/2020	30°58’6.10” E	25°26’37.92” S
	C. quilicii	Pineapple guava	(*Feijoa sellowiana* (O.Berg) O.Berg)	05/2021	29°59’ 11.56” E	26°30’51.31” S
	C. rosa	Jambos	(*Syzygium jambos* L. Alston)	11/27/2020	30°58’10.99’’ E	25°27’08.54’’ S
	B. dorsalis	Mango	(*Mangifera indica* L.)	2/12/2021	30°57’ 15.84” E	25°32’58.06” S

Table S2. Sample collection data for the wild, trap-collected specimens used for assay validation in this study. The collection site is provided as the province and coordinates.

Species	Collection date	Province	Coordinates
B. dorsalis	08/2021	Mpumalanga	31°04'17.41" E 25°26'38.27" S
	07/2021	Mpumalanga	30°34'31.48" E 25°23'52.07" S
	06/2021	Limpopo	30°50'51.95" E 24°24'26.61" S
	09/2021	Limpopo	30°32'36.02" E 23°45'18.95" S
	06/2021	Limpopo	30°22'54.60" E 23°52'21.32" S
C. quilicii	06/2021	Northwest	25°46'12.9" S 27°36'51.7" E
	06/2021	Mpumalanga	25°22'50.7" S 30°32'02.2" E
	06/2021	KwaZulu Natal	27°21'44.7" S 31°47'14.4" E
	03/2021	Free State	28°15'09.1" S 28°19'02" E
	06/2021	Eastern Cape	33°36'43.4" S 25°39'39.1" E
C. capitata	06/2021	Northwest	25°46'12.9" S 27°36'51.7" E
	06/2021	Limpopo	23°51'47.7" S 30°23'08.4" E
	06/2021	Mpumalanga	25°26'39.3" S 31°33'15.4" E
	06/2021	KwaZulu Natal	27°21'44.7" S 31°47'14.4" E
	05/2021	Northern Cape	28°48'8.83" S 20°39'56.2" E
C. cosyra	06/2021	Northwest	25°46'12.9" S 27°36'51.7" E
	06/2021	Limpopo	23°51'47.7" S 30°23'08.4" E
	06/2021	Mpumalanga	25°24'34.3" S 30°55'46.6" E
	08/2021	KwaZulu Natal	27°21'44.7" S 31°47'14.4" E
C. rosa	06/2021	Mpumalanga	25°26'44.9" S 30°58'05.1" E

Table S3. List of complete mitochondrial genomes used for primer design and multiple sequence comparison.

Species	Accession number
Ceratitis quilicii	MT998948.1
	MT036776.1
	MT036777.1
	MT036778.1
	MT036779.1
	MT036780.1
	MT036791.1
	MT036792.1
	MT036793.1
	MT036794.1
	MT036795.1
	MT036790.1
	NC_053846.1
Ceratitis rosa	MT036796.1
	MT036797.1
	MT036798.1
	MT036799.1
	MT036800.1
	MT997010.1
Ceratitis capitata	AJ242872.1
	NC_000857.1
	MT036782.1
Ceratitis cosyra	MT036783.1
	MT036784.1
Bactrocera dorsalis	KT343905.1
	DQ917577.1
	DQ845759.1
	NC_008748.1
	MG916968.1
	MN104220.1

Table S4. BLASTn results showing a high percentage identity between query sequences and members of the same species for each mitogenome generated in this study.

Specimen	Hit	Accession	E-value	% Identity
C. capitata (1)	*Ceratitis capitata*	NC_000857.1	0.0	99.74
C. capitata (2)	*Ceratitis capitata*	NC_000857.1	0.0	99.74
C. cosyra (1)	*Ceratitis cosyra*	MT036784.1	0.0	99.50
C. cosyra (3)	*Ceratitis cosyra*	MT036784.1	0.0	99.20
C. rosa (1)	*Ceratitis rosa*	MT036799.1	0.0	99.87
C. rosa (2)	*Ceratitis rosa*	MT036799.1	0.0	99.87
C. quilicii (1)	*Ceratitis quilicii*	MT036776.1	0.0	99.91
C. quilicii (2)	*Ceratitis quilicii*	MT036778.1	0.0	99.67
B. dorsalis (1)	*Bactrocera dorsalis*	MN104220.1	0.0	99.86
B. dorsalis (2)	*Bactrocera dorsalis*	MN104220.1	0.0	99.87

APPENDIX B. SUPPLEMENTARY MATERIAL FOR CHAPTER 4

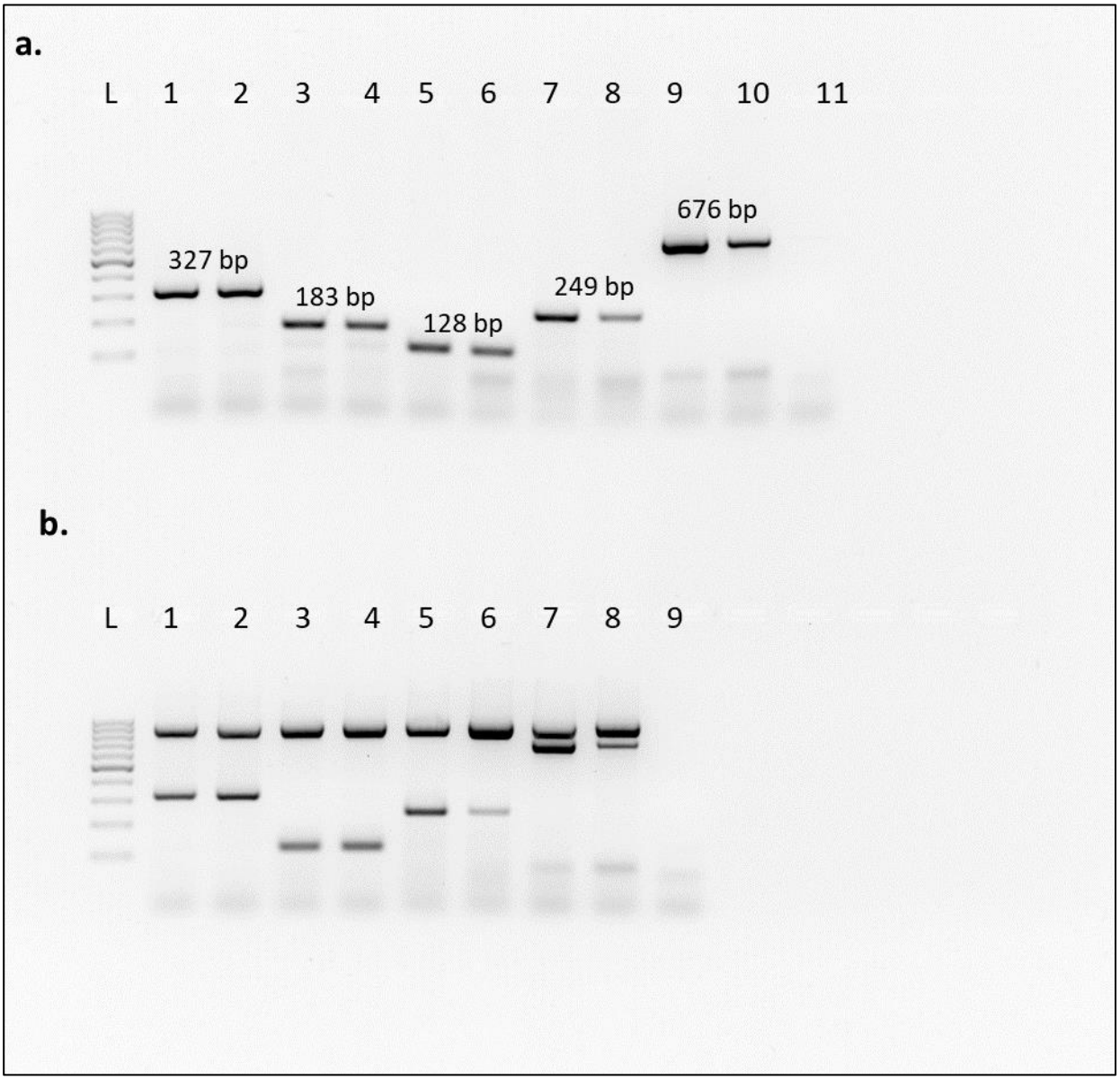


Figure S1. (a.) 2% agarose-TAE gel displaying the specificity of the multiplex PCR assay on freshly extracted DNA from colony-reared insects with species-specific amplicon size indicated. Lanes 1 & 2: *C. capitata*, Lanes 3 & 4: *C. cosyra*, Lanes 5 & 6: *C. quilicii*, Lanes 7 & 8: *C. rosa*, Lanes 9 & 10: *B. dorsalis*, Lane 11: No template control, Lane L: GeneRuler 100 bp DNA ladder (Thermo Scientific). **(b.)** 2% agarose-TAE gel displaying the multiplex PCR (excluding *C. cosyra* primers) in conjunction with the universal primer set CI-J2183 and TL2-N3014 amplifying the COI gene in colony reared insects. Lanes 1 & 2: *C. capitata*, Lanes 3 & 4: *C. quilicii*, Lanes 5 & 6: *C. rosa*, Lanes 7 & 8: *B. dorsalis*, Lane 9: No template control, Lane L: GeneRuler 100 bp DNA ladder (Thermo Scientific).

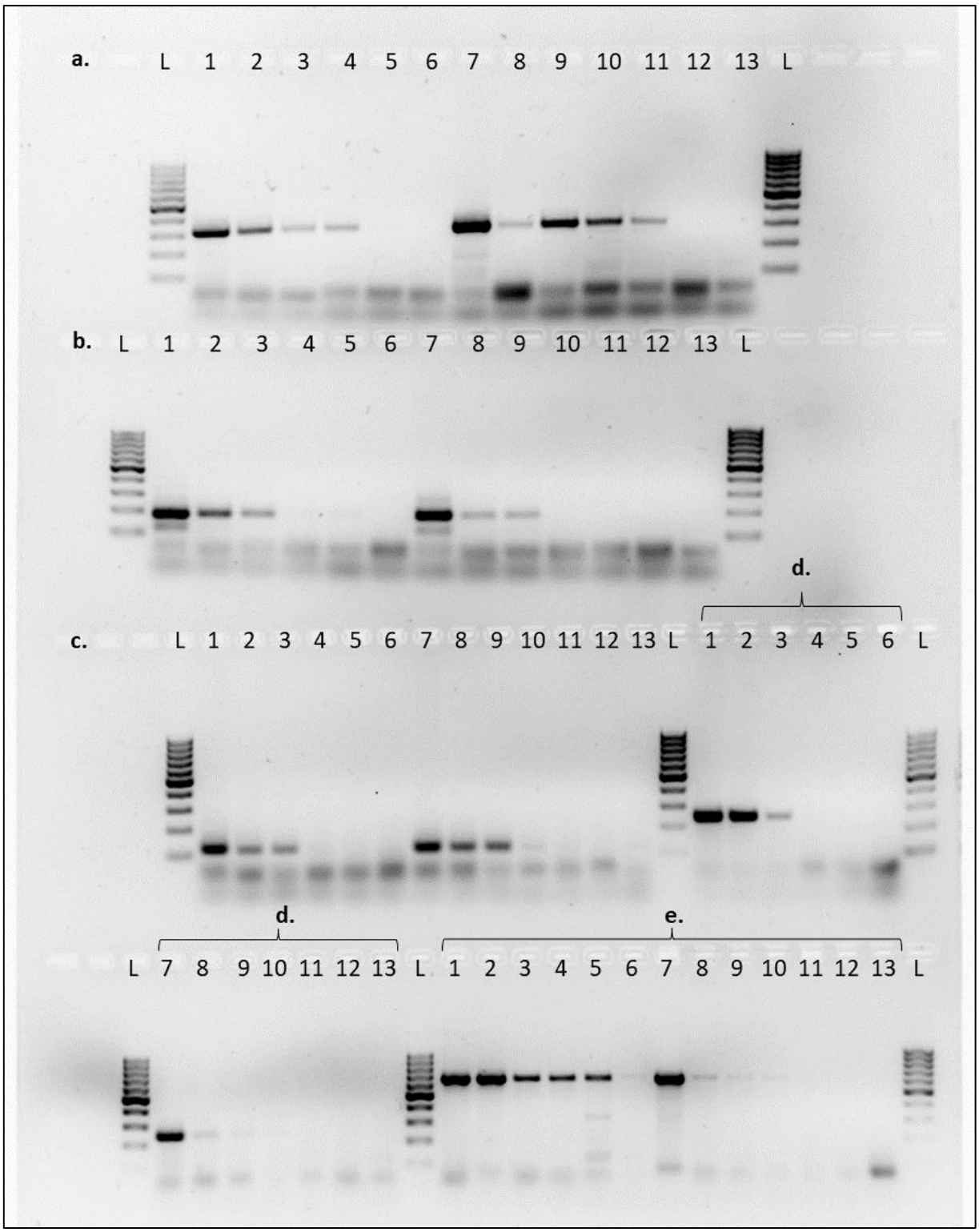


Figure S2. 2% agarose TAE gel stained with ethidium bromide displaying the sensitivity of the multiplex PCR assay on adult male colony specimens. Row **(a.)** contains *C. capitata*, row **(b.)** contains *C. cosyra*, row **(c.)** contains *C. quilicii*, row **(d.)** contains *C. rosa*, and row **(e.)** contains *B. dorsalis*. Lanes labelled (L) contain 100 bp DNA ladder (Thermo Scientific), lanes labelled (1 – 6) correspond to varying concentrations in the 5x dilution series for the first DNA sample, lanes labelled (7 - 12) correspond to the same concentrations repeated for a second DNA sample of the same species, and lanes labelled (13) contain a no template control (NTC). Lane 1 & 7: 20 ng/µl, lane 2 & 8: 4 ng/µl, lane 3 & 9: 0.8 ng/µl, lane 4 & 10: 0.16 ng/µl, lane 5 & 11: 0.32 ng/µl, and lane 6 & 12: 0.0064 ng/µl.

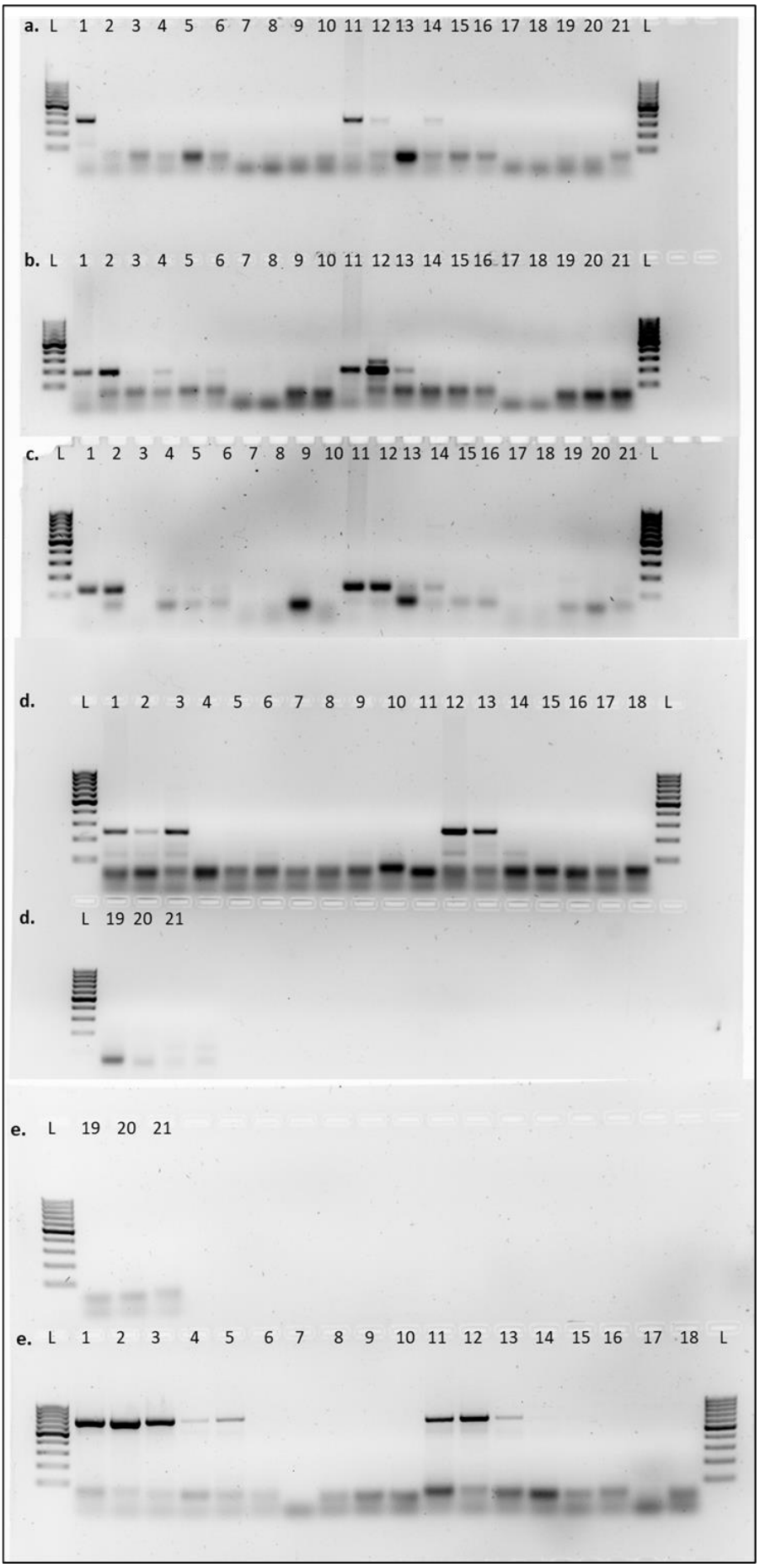


Figure S3. 2% agarose TAE gel stained with ethidium bromide displaying the sensitivity of the multiplex PCR assay on larval colony specimens. Row **(a.)** contains *C. capitata*, row **(b.)** contains *C. cosyra*, row **(c.)** contains *C. quilicii*, row **(d.)** contains *C. rosa*, and row **(e.)** contains *B. dorsalis*. Lanes labelled (L) contain 100 bp DNA ladder (Thermo Scientific), lanes labelled (1 – 10) correspond to varying concentrations in the 5x dilution series for the first DNA sample, lanes labelled (11 - 20) correspond to the same concentrations repeated for a second DNA sample of the same species, and lanes labelled (21) contain a no template control (NTC). Lane 1 & 11: 50 ng/µl, lane 2 & 12: 10 ng/µl, lane 3 & 13: 2 ng/µl, lane 4 & 14: 0.4 ng/µl, lane 5 & 15: 0.08 ng/µl, lane 6 & 16: 0.0016 ng/µl, lane 7 & 17: 0.0032 ng/µl, lane 8 & 18: 0.00064 ng/µl, lane 9 & 19: 0.000128 ng/µl, and lane 10 & 20: 0.0000256 ng/µl.

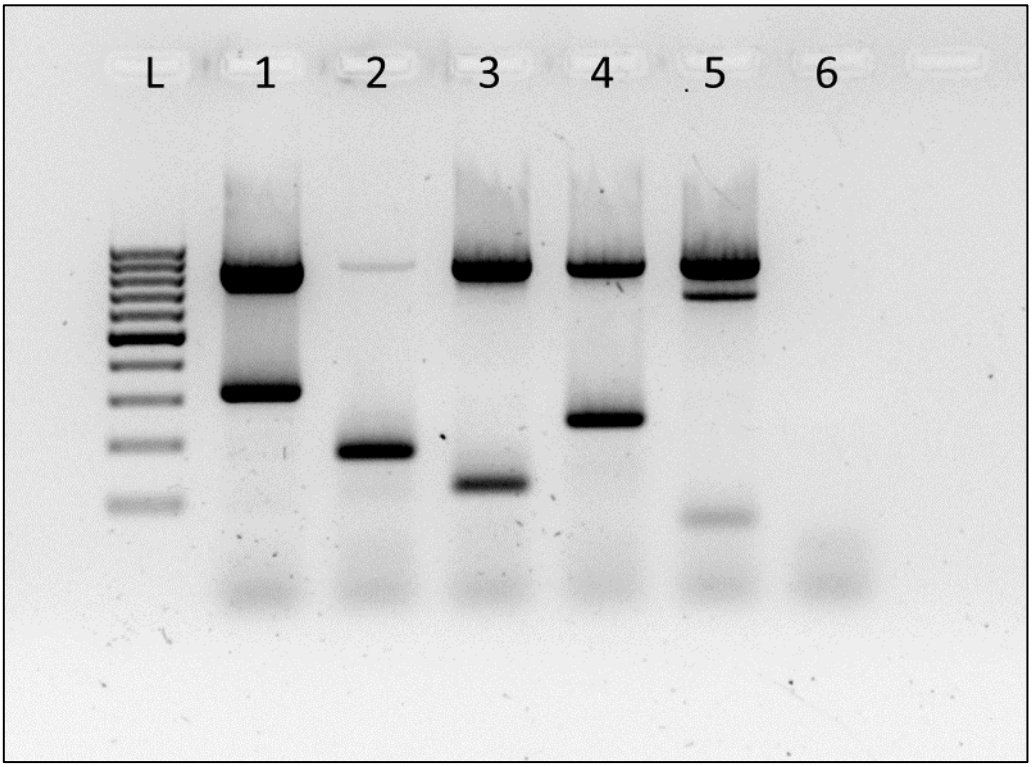


Figure S4. 2% agarose TAE gel stained with ethidium bromide displaying results of multiplex PCR on colony adult males including the COI internal control. Lane 1: *C. capitata*, lane 2: *C. cosyra*, Lane 3: *C. quilicii*, Lane 4: *C. rosa*, Lane 5: *B. dorsalis*, Lane 6: no template control (NTC), and lane L: 100 bp DNA ladder (Thermo Scientific).

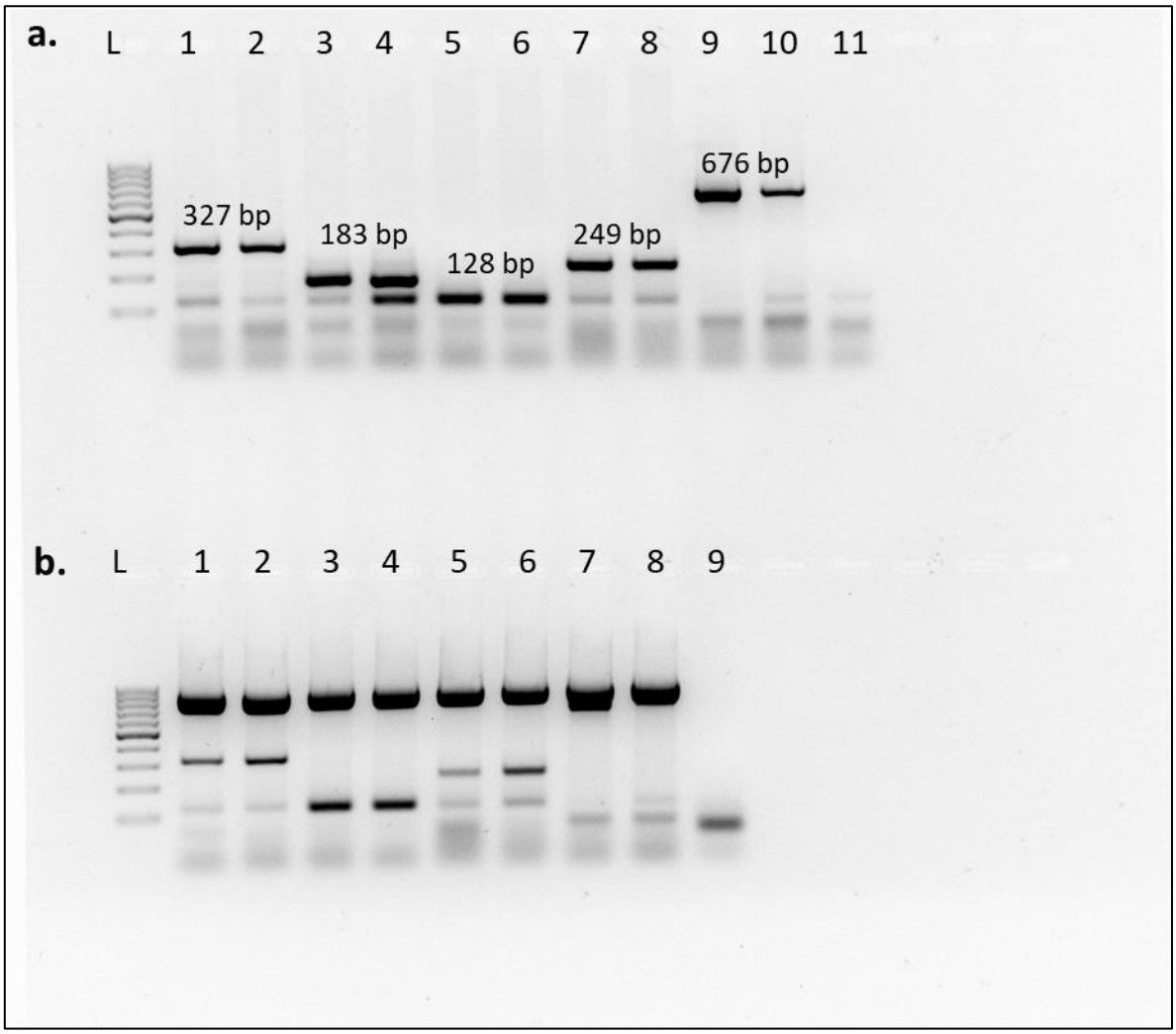


Figure S5 (a.). 2% Agarose-TAE gel displaying the specificity of multiplex primers in the case of duplex formation in freshly extracted colony-reared larval DNA with species-specific amplicon size indicated. Lane 4 demonstrates the expected *C. cosyra* amplicon at 183 bp with non-specific amplification at 128 bp leading to the formation of a duplex; the larger 183 bp amplicon should be used for identification. Lane 5 is a single 128 bp amplicon indicative of *C. quilicii*. Lanes 1 & 2: *C. capitata*, Lanes 3 & 4: *C. cosyra*, Lanes 5 & 6: *C. quilicii*, Lanes 7 & 8: *C. rosa*, Lanes 9 & 10: *B. dorsalis*, Lane 11: No template control, Lane L: GeneRuler 100 bp DNA ladder (Thermo Scientific). **(b.)** 2% Agarose-TAE gel displaying the multiplex PCR (excluding *C. cosyra* primers) in conjunction with the universal primer set CI-J2183 and TL2-N3014 amplifying the COI gene in colony reared larvae. Lanes 1 & 2: *C. capitata*, Lanes 3 & 4: *C. quilicii*, Lanes 5 & 6: *C. rosa*, Lanes 7 & 8: *B. dorsalis*, Lane 9: No template control, Lane L: GeneRuler 100 bp DNA ladder (Thermo Scientific).

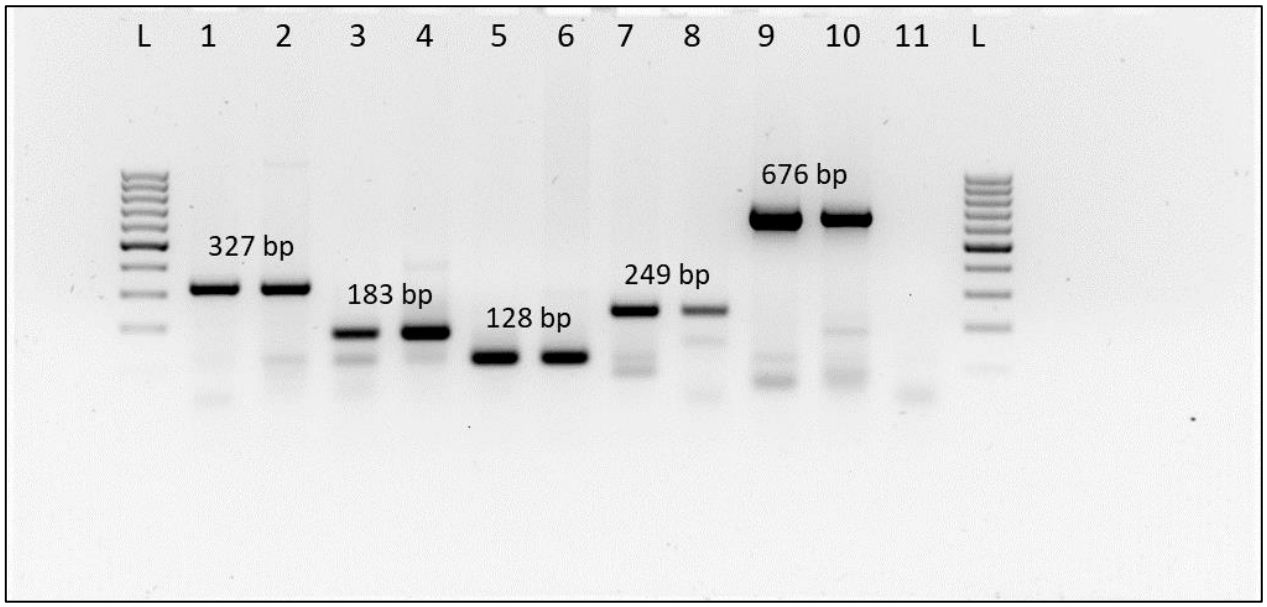


Figure S6. 2% agarose-TAE gel displaying the efficacy of the Multiplex PCR assay to identify wild, trap-collected specimens with examples of non-specific amplification. Species-specific amplicon sizes are indicated. Lanes 1 & 2: *C. capitata*, Lanes 3 & 4: *C. cosyra*, Lanes 5 & 6: *C. quilicii*, Lanes 7 & 8: *C. rosa*, Lanes 9 & 10: *B. dorsalis*, Lane 11: No template control, Lane L: GeneRuler 100 bp DNA ladder (Thermo Scientific).

Table S1. Origin of colony flies reared at the CRI (Mbombela, Mpumalanga, South Africa). Adult colony insects collected in February 2021 were refreshed between January 2019 and January 2020. Larval specimens collected in August 2021 were refreshed between November 2020 and May 2021. Collection sites are provided as coordinates.

Life stage	Species	Rearing fruit		Collection Date	Latitude	Longitude
Adult male	*C. capitata*	Coffee	(*Coffea canephora Pierre ex. Froehner)*	10/03/2019	31°5'15.63" E	25°6'43.50" S
	C. cosyra	Marula	(*Sclerocarya birrea* (A. Rich.) Hochst.)	1/29/2019	31°2'35.70" E	25°28'4.03" S
	C. quilicii	Peach	(*Prunus persica* L. Batsch)	1/21/2020	30°23'34.73" E	24°59'47.17" S
	C. rosa	Strawberry Guava	(*Psidium cattleyanum* Sabine)	11/13/2019	30°58'10.99'' E	25°27'08.54'' S
	B. dorsalis	Mango	(*Mangifera indica* L.)	1/29/2019	30°58'10.99'' E	25°27'08.54'' S
Larvae	*C. capitata*	Coffee	(*Coffea canephora Pierre ex.* Froehner)	3/10/2021	31°5'15.63" E	25°6'43.50" S
	C. cosyra	Pepper-bark tree	(*Warburgia salutaris* (Bertol.f.) Chiov.)	11/30/2020	30°58'6.10" E	25°26'37.92" S
	C. quilicii	Pineapple guava	(*Feijoa sellowiana* (O.Berg) O.Berg)	05/2021	29°59' 11.56" E	26°30'51.31" S
	C. rosa	Jambos	(*Syzygium jambos* L. Alston)	11/27/2020	30°58'10.99'' E	25°27'08.54'' S
	B. dorsalis	Mango	(*Mangifera indica* L.)	2/12/2021	30°57' 15.84" E	25°32'58.06" S

Table S2. List of gene regions utilised in the preliminary bioinformatic analysis, including accession numbers of the reference sequences and the literature from which they were selected.

Gene region investigated	Reference sequence *C. capitata*	Reference sequence *B. dorsalis*	Literature cited
Cytochrome c oxidase subunit I	MW410928.1	MZ621836.1	(Armstrong and Ball, 2005; Garzón-Orduña *et al.*, 2020)
Cytochrome c oxidase subunit II	NC_000857.1	KT343905.1	(Armstrong and Ball, 2005; Garzón-Orduña *et al.*, 2020)
Protein MTO1 homolog	XM_004517805.4	XM_011208269.2	Not available
Glutamine synthetase 1, mitochondrial	XM_004530302.3	XM_011209495.2	(Killer *et al.*, 2020)
Presequence protease, mitochondrial	XM_004535774.4	XM_019991906.1	(Alikhani *et al.*, 2011)
ERF3A	GAMC01016820.1	GAKP01021067.1	(Tarasov, Zhuravleva and Abramson, 2008)
Dynamin	GAMC01011620.1	XM_011213658.3	(Hardy, 2007)
TipE	XM_012302544.2	XM_019992132.1	(Bourdin *et al.*, 2015)
Gustatory receptor for bitter taste 22e-like	XM_023303386.1	XM_011212444.1	(Papanicolaou *et al.*, 2016)
Gustatory and pheromone receptor 32a	XM_004526635.3	XM_019991446.1	
CCR4-NOT transcription complex subunit 6-like	XM_012300675.2	XM_019990104.1	
Toll-like receptor Tollo	XM_004522201.2	XM_011209703.2	
Gawky	XM_020860629.1	XM_011203056.2	
Opsin Rh2	XM_004534311.3	XM_011205152.2	
Opsin Blue Sensitive	XM_004525726.3	N/A	
Opsin Rh4	XM_004526176.3	XM_011215866.2	
Opsin Rh1	XM_004535527.3	XM_018947745.1	
Opsin Rh6	XM_004518077.2	XM_011209950.2	

References:

Alikhani, N. *et al.* (2011) 'Targeting Capacity and Conservation of PreP Homologues Localization in Mitochondria of Different Species', *Journal of Molecular Biology*, 410(3), pp. 400–410. doi: 10.1016/j.jmb.2011.05.009.

Armstrong, K. F. and Ball, S. L. (2005) 'DNA barcodes for biosecurity: invasive species identification', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1462), pp. 1813–1823. doi: 10.1098/rstb.2005.1713.

Bourdin, C. M. *et al.* (2015) 'Molecular and functional characterization of a novel sodium channel TipE-like auxiliary subunit from the American cockroach *Periplaneta americana*', *Insect Biochemistry and Molecular Biology*, 66, pp. 136–144. doi: 10.1016/j.ibmb.2015.10.008.

Garzón-Orduña, I. J. *et al.* (2020) 'Implementing Low-Cost, High Accuracy DNA Barcoding From Single Molecule Sequencing to Screen Larval Tephritid Fruit Flies Intercepted at Ports of Entry', *Annals of the Entomological Society of America*. Edited by K. Gaddis, 113(4), pp. 288–297. doi: 10.1093/aesa/saz071.

Hardy, N. B. (2007) 'Phylogenetic utility of dynamin and triose phosphate isomerase', *Systematic Entomology*, 32(2), pp. 396–403. doi: 10.1111/j.1365-3113.2007.00377.x.

Killer, J. *et al.* (2020) 'Glutamine synthetase type I (glnAI) represents a rewarding molecular marker in the classification of bifidobacteria and related genera', *Folia Microbiologica*, 65(1), pp. 143–151. doi: 10.1007/s12223-019-00716-0.

Papanicolaou, A. *et al.* (2016) 'The whole genome sequence of the Mediterranean fruit fly, *Ceratitis capitata* (Wiedemann), reveals insights into the biology and adaptive evolution of a highly invasive pest species', *Genome Biology*, 17(1), p. 192. doi: 10.1186/s13059-016-1049-2.

Tarasov, O. V., Zhuravleva, G. A. and Abramson, N. I. (2008) 'Evaluation of the gene encoding translation termination factor eRF3 as a possible phylogenetic marker', *Molecular Biology*, 42(6), pp. 834–842. doi: 10.1134/S0026893308060022.

Table S3. List of *Wolbachia* reference sequences used for local BLASTn database analysis of de novo assembled contigs.

Accession number:	Reference:
CP041924	Not available
AB036666.1	(Masui *et al.*, 2000)
AB478515.1	(Tanaka *et al.*, 2009)
AE017196.1	(Wu *et al.*, 2004)
AM999887.1	(Klasson *et al.*, 2008)
CP001391.1	(Klasson *et al.*, 2009)
CP003883.1	(Ellegaard *et al.*, 2013)
CP011148.1	Not available
CP015510.2	(Faddeeva-Vakhrusheva *et al.*, 2017)
HQ906664.1	(Kent *et al.*, 2011)

References:

Ellegaard, K. M. *et al.* (2013) 'Comparative Genomics of *Wolbachia* and the Bacterial Species Concept', *PLoS Genetics*. Edited by I. Matic, 9(4), p. e1003381. doi: 10.1371/journal.pgen.1003381.

Faddeeva-Vakhrusheva, A. *et al.* (2017) 'Coping with living in the soil: the genome of the parthenogenetic springtail *Folsomia candida*', *BMC Genomics*, 18(1), p. 493. doi: 10.1186/s12864-017-3852-x.

Kent, B. N. *et al.* (2011) 'Complete Bacteriophage Transfer in a Bacterial Endosymbiont (*Wolbachia*) Determined by Targeted Genome Capture', *Genome Biology and Evolution*, 3, pp. 209–218. doi: 10.1093/gbe/evr007.

Klasson, L. *et al.* (2008) 'Genome Evolution of *Wolbachia* Strain wPip from the Culex pipiens Group', *Molecular Biology and Evolution*, 25(9), pp. 1877–1887. doi: 10.1093/molbev/msn133.

Klasson, L. *et al.* (2009) 'The mosaic genome structure of the *Wolbachia* w Ri strain infecting *Drosophila simulans*', *Proceedings of the National Academy of Sciences*, 106(14), pp. 5725–5730. doi: 10.1073/pnas.0810753106.

Masui, S. *et al.* (2000) 'Distribution and Evolution of Bacteriophage WO in *Wolbachia*, the Endosymbiont Causing Sexual Alterations in Arthropods', *Journal of Molecular Evolution*, 51(5), pp. 491–497. doi: 10.1007/s002390010112.

Tanaka, K. *et al.* (2009) 'Complete WO Phage Sequences Reveal Their Dynamic Evolutionary Trajectories and Putative Functional Elements Required for Integration into the *Wolbachia* Genome', *Applied and Environmental Microbiology*, 75(17), pp. 5676–5686. doi: 10.1128/AEM.01172-09.

Wu, M. *et al.* (2004) 'Phylogenomics of the Reproductive Parasite *Wolbachia pipientis* wMel: A Streamlined Genome Overrun by Mobile Genetic Elements', *PLoS Biology*. Edited by Nancy A. Moran, 2(3), p. e69. doi: 10.1371/journal.pbio.0020069.

Table S4. Sample collection data for the wild, trap-collected specimens used for assay validation in this study. The collection site is provided as the province and coordinates.

Species	Collection date	Province	Coordinates
B. dorsalis	08/2021	Mpumalanga	31°04'17.41" E 25°26'38.27" S
	07/2021	Mpumalanga	30°34'31.48" E 25°23'52.07" S
	06/2021	Limpopo	30°50'51.95" E 24°24'26.61" S
	09/2021	Limpopo	30°32'36.02" E 23°45'18.95" S
	06/2021	Limpopo	30°22'54.60" E 23°52'21.32" S
C. quilicii	06/2021	Northwest	25°46'12.9" S 27°36'51.7" E
	06/2021	Mpumalanga	25°22'50.7" S 30°32'02.2" E
	06/2021	KwaZulu Natal	27°21'44.7" S 31°47' 14.4" E
	03/2021	Free State	28°15'09.1" S 28°19'02" E
	06/2021	Eastern Cape	33°36'43.4" S 25°39'39.1" E
C. capitata	06/2021	Northwest	25°46'12.9" S 27°36'51.7" E
	06/2021	Limpopo	23°51'47.7" S 30°23'08.4" E
	06/2021	Mpumalanga	25°26'39.3" S 31°33'15.4" E
	06/2021	KwaZulu Natal	27°21'44.7" S 31°47' 14.4" E
	05/2021	Northern Cape	28°48'8.83" S 20°39'56.2" E
C. cosyra	06/2021	Northwest	25°46'12.9" S 27°36'51.7" E
	06/2021	Limpopo	23°51'47.7" S 30°23'08.4" E
	06/2021	Mpumalanga	25°24'34.3" S 30°55'46.6" E
	08/2021	KwaZulu Natal	27°21'44.7" S 31°47'14.4" E
C. rosa	06/2021	Mpumalanga	25°26'44.9" S 30°58'05.1" E

APPENDIX C. CONFERENCE POSTER

A new diagnostic tool for the identification of fruit fly larvae in citrus

KJ ANDREWS[1], R BESTER[1,2], A MANRAKHAN[3,4], HJ MAREE[1,2]

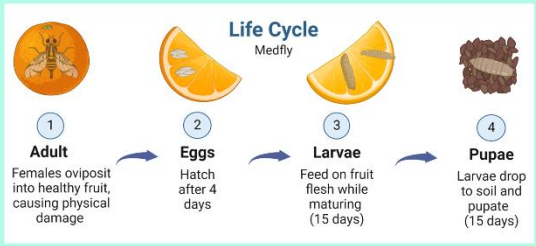
[1]Department of Genetics, Stellenbosch University, Private bag X1, Matieland 7602, South Africa
[2]Citrus Research International, PO Box 2201, Matieland 7602, South Africa
[3]Citrus Research International, PO Box 28, Mbombela 1200, South Africa
[4]Department of Conservation Ecology and Entomology, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa

BACKGROUND

South Africa is afflicted with five fruit flies of economic concern which cause severe damage to fruit while completing their life cycle[1]. These fruit flies are frugivorous (attacking fresh fruit), polyphagous[2] (having multiple hosts), and potentially invasive[3], a major risk for importing countries of fresh fruit from SA. This has resulted in the restriction and limitations of SA export markets[4]. These fruit flies are morphologically very similar, posing an issue for accurate morphological identification, especially in early life stages[5]. In the case of phytosanitary pests, which includes the five fruit flies under study, accurate identification is essential for pest surveillance and successful management practices, as well as for quarantine decisions made at Ports of Entry (PoE).

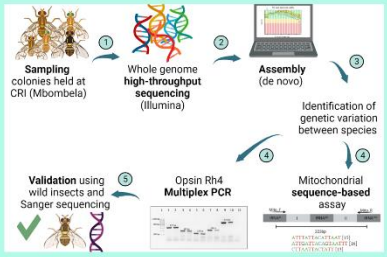
OBJECTIVE

The following study set out to develop a molecular identification assay to quickly and accurately identify these five fruit flies to species level regardless of the life stage intercepted.



Ceratitis Capitata (Med fly) | *Ceratitis Cosyra* (Marula fly) | *Ceratitis rosa* (Natal fly) | *Ceratitis quilicii* (Cape fly) | *Bactrocera dorsalis* (Oriental fly)

METHODS & RESULTS



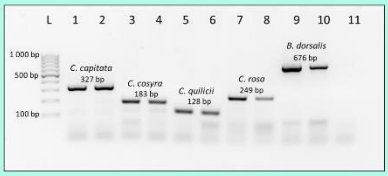
Two identification assays were developed in this study:

Sequencing Based Assay

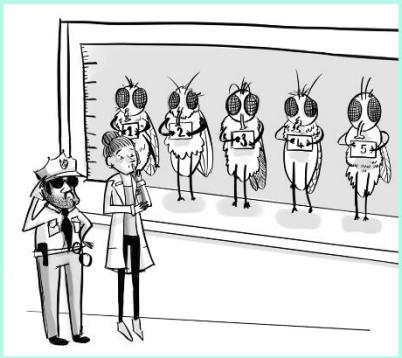
- Targets a mitochondrial intergenic region
- Takes ± 3 – 4 days for results
- Best suited for pest surveillance

Multiplex PCR

- Targets the gene Opsin Rh4
- Takes ± 2.5 hours for results (post DNA extraction)
- Best suited for time-sensitive cases such as interceptions at PoE



2% Agarose TAE gel stained with ethidium bromide displaying results of the Multiplex PCR. Lanes 1&2: *C. capitata*, Lanes 3&4: *C. cosyra*, Lanes 5&6: *C. quilicii*, Lanes 7&8: *C. rosa*, Lanes 9&10: *B. dorsalis*, Lane 11: No template control, Lane L: 100 bp DNA ladder



CONCLUSION

The molecular assays designed in this study accurately and reliably differentiate between five fruit fly species of economic importance to South Africa. These assays can be used in a stand-alone format or in conjunction with morphological identification for successful monitoring and management practices in addition to assisting decisions of quarantine significance at PoE.

Scan me for access to research article

Scan me for a copy of this poster

ACKNOWLEDGEMENTS

We would like to acknowledge **Citrus Research International** (CRI) for funding project 1299 and extend thanks to **Stellenbosch University** for use of their facilities, and **Ilani Mostert** for her illustrations.

References:

[1]White, I. M. and Elson-Harris, M. M. (1992) *Fruit flies of economic significance: their identification and bionomics*. CAB international.
[2]Virgilio, M., White, I. and De Meyer, M. (2014) 'A set of multi-entry identification keys to African frugivorous flies (Diptera, Tephritidae)', *ZooKeys*, 428, pp. 97–108.
[3]Malacrida, A. R. *et al.* (2007) 'Globalization and fruitfly invasion and expansion: the medfly paradigm', *Genetica*, 131(1), pp. 1–9.
[4]Louzeiro, L. R. F. *et al.* (2021) 'Incidence of frugivorous flies (Tephritidae and Lonchaeidae), fruit losses and the dispersal of flies through the transportation of fresh fruit', *Journal of Asia-Pacific Entomology*, 24(1), pp. 50–60.
[5]Boykin, L. M. *et al.* (2012) 'Species Delimitation and Global Biosecurity', *Evolutionary Bioinformatics*, 8, p. EBO.S8532.
[6]Images adapted from the Royal Museum for Central Africa online collection. ©Africa Museum, image author: Jonathan Brecko.

www.nature.com/scientificreports

scientific reports

Check for updates

OPEN

A multiplex PCR assay for the identification of fruit flies (Diptera: Tephritidae) of economic importance in South Africa

Kelsey J. Andrews[1], Rachelle Bester[1,2], Aruna Manrakhan[3,4] & Hans J. Maree[1,2]✉

The fruit fly (Diptera: Tephritidae) species, *Ceratitis capitata*, *Ceratitis cosyra*, *Ceratitis rosa*, *Ceratitis quilicii*, and *Bactrocera dorsalis* are of economic importance in South Africa. These agricultural pests cause extensive damage to a range of commercially produced fruit, primarily for export. These pests are of phytosanitary significance, and their presence in fruit-producing regions in South Africa has led to restrictions in export trade of fresh produce. Accurate identification of these flies, particularly at immature stages intercepted in fruit consignments originating from South Africa, is essential but remains an ongoing challenge. A rapid and accurate identification assay to differentiate these five species is needed for inspection and pest surveillance. High throughput sequencing data were generated for each of the five fruit fly species, and five sets of species-specific primers were designed for use in a multiplex PCR. Each primer set amplifies an amplicon of a different size for each species allowing for accurate identification. PCR sensitivity tests demonstrate that the limit of detection for this assay is 10 ng and 4 ng of DNA when extracted from larvae and adult specimens, respectively. The assay developed can be applied in fruit inspection and survey activities within the country and at ports of entry.

Tephritidae is an agriculturally important family with many fruit fly species known to cause extensive damage to commercial fruit[1]. Quarantine restrictions are in place to limit any further spread of these fruit fly pests. In South Africa, five economically important fruit flies are present that can potentially affect the production and export of commercial fruit[2–4]. They are *C. capitata* (Wiedemann), Mediterranean fruit fly; *C. cosyra* (Walker), marula fly; *C. rosa* (Karsch), Natal fly; *C. quilicii* De Meyer, Mwatawala & Virgilio, Cape fly; and *B. dorsalis* (Hendel), the Oriental fruit fly. *Ceratitis quilicii* is a recently described species[3], hence the pest status and host range of this species in commercial fruit production areas in South Africa are still being determined. The *Ceratitis* species are of Afrotropical origin[5] while *B. dorsalis* is of Asian origin and was introduced in the northeastern parts of South Africa in 2013[4]. The five fruit fly species are polyphagous (attacking fruit from different plant families)[6], and two of them, *C. capitata* and *B. dorsalis*, have demonstrated a high affinity for invasiveness with significant expansion of their distribution beyond their native ranges[7,8]. This is a major challenge for horticultural and export industries, particularly with the increasing frequency of international trade[9]. These five fruit flies are currently the only major tephritid pests of commercial fresh fruit produced primarily for export from South Africa.

South Africa is a significant producer and exporter of fresh fruit. In the 2019/2020 production season, over 6.5 million metric tons of fruit were produced, and more than half of the total produce was exported (Fruit South Africa, *2020 Key Fruit Statistics*). Fruit fly pests are of phytosanitary significance for fresh fruit exported from South Africa. It is not uncommon for multiple fruit fly species to infest the same commercial fruit, as the host range of these fruit flies often overlap[10]. South African fruit must meet the country-specific phytosanitary requirements of the export markets to prevent the entry of fruit fly pests[11]. The interception of phytosanitary pests on consignments at Ports of Entry (PoE) can result in the destruction of the commodity or return of the commodity to the country of origin[12]. The time required to accurately identify any pests present in consignments delays the shipment of fresh produce. Fresh fruit and vegetables may be detained for days while undergoing inspection, reducing their economic value. The European Union (EU), an important export market for fresh

[1]Department of Genetics, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa. [2]Citrus Research International, PO Box 2201, Matieland 7602, South Africa. [3]Citrus Research International, PO Box 28, Mbombela 1200, South Africa. [4]Department of Conservation Ecology and Entomology, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa. ✉email: hjmaree@sun.ac.za

fruit from South Africa, has zero-tolerance enforcement for non-EU Tephritidae, including all fruit fly pests in South Africa, except *C. capitata*, which is an established pest in the EU[13]. There is a need to be able to distinguish between *C. capitata* and the other four fruit fly pests during inspection before and after export for markets such as the EU. Therefore, a rapid method to accurately identify the five fruit fly pests infesting fresh fruit in South Africa is essential.

Inspection and survey of fruit fly species are often largely reliant on morphological identification of specimens by expert taxonomists and published keys[1,6,14]. The morphological identification of fruit flies to species level can be more reliably made at the adult stage, either emerged adults from infested fruit or adults collected from traps, using these keys. The difficulty arises in differentiating between cryptic species or damaged adult specimens where few distinguishing morphological differences exist, and female specimens appear near identical. However, when immature stages are intercepted, either eggs or larvae in fruit or pupae in soil, and development to adulthood is not practical due to time sensitivity, identification to species level using molecular methods would be more appropriate[15]. There are keys to differentiate between species at the larval stages. Still, these can only be used at the third instar stage and remain problematic if either the specimen is not in good condition, earlier life stages are intercepted, or pupae are found. Larval identification using these keys does not enable the identification of closely related species or species complexes[14]. *Ceratitis rosa* and *C. quilicii* are examples of cryptic species that were previously thought to be the same species until their recent separation in 2016 based on morphological and genetic differences[3,16]. Research has demonstrated that the taxonomic classification of many of these cryptic fruit fly species cannot reliably be resolved through morphological characterization alone, where population-level variation can be easily confused with species-level variation[17]. A molecular-based identification assay may alleviate some of the difficulties experienced in the morphological identification of such closely related species.

Molecular identification tools can offer the advantage of a faster turnaround time as the life stage of the specimen is not a limiting factor. Broad detection assays for tephritid fruit flies have been developed, although they do not reliably allow for identification to species level[18,19]. Microsatellite markers have been considered for identification of closely related fruit fly species, however, this can be expensive and time consuming where six to 16 markers have to be used and unambiguous species identification is not possible without prior morphological identification[20–22]. Molecular identification techniques for fruit flies have been primarily centered around DNA barcoding using cytochrome c oxidase subunit I (COI). Although this method can differentiate between many fruit fly species, it cannot accurately differentiate between species complexes such as FARQ (*C. fasciventris* (Bezzi), *C. anonae* Graham, *C. rosa*, and *C. quilicii)*[3,23,24] and the *B. dorsalis* complex[25]. Literature suggests that the failure of COI to identify closely related species may be due to incomplete lineage sorting within these species complexes[25]. Misidentifications can be reduced by introducing a distance threshold, where a query sequence is considered unidentifiable if the closest DNA barcode match exceeds the value of the distance threshold set. However, if the distance threshold is too restrictive, it is at the sacrifice of reduced barcoding accuracy with a higher proportion of discarded queries[26]. DNA barcoding relies on time-consuming DNA sequencing, an additional expenditure not applicable for routine analysis[27,28]. COI has also been used for qPCR and real-time PCR identification assays[18,29]. Expansions into other mitochondrial genes for PCR–RFLP analysis and Tephritidae identification have also been explored[27,30]. However, most molecular identification assays based on mitochondrial genes had limitations in identifying closely related species and species complexes.

An investigation into the use of genomic regions as opposed to mitochondrial research for tephritid fruit fly identification was undertaken in this study. The ability to identify multiple species simultaneously and rapidly without the need for costly downstream analysis and sequencing was deemed a priority. Multiplex PCR offers the ability to amplify different DNA targets and different amplicon sizes in a single run. Although the use of multiplex PCR for fruit fly identification has not been well explored, it has shown promising results in differentiating a species of interest, *Rhagoletis cerasi* Loew, from other tephritid flies in North America as well as fruit fly parasitoid identification[31,32]. While the five fruit flies under study can be identified through a variety of existing molecular assays, to date, no assay can identify all five flies simultaneously. Therefore, this study utilizes a multiplex PCR approach to provide a fast and accurate identification assay for differentiation of five tephritid fruit flies of economic importance to South Africa without life-stage restrictions. Many of these tephritid fruit fly species also occur in other parts of Africa[3,5,21,33]. As such the development of a rapid and accurate identification technique in this study will be applicable for fruit fly identification in other parts of Africa where these species occur and are of economic importance.

Results

Species identification and DNA extraction. All adult specimens used in this study underwent morphological identification and DNA barcoding targeting the COI gene with the universal primer pair CI-J2183 and TL2-N3014[34]. All adult specimens were identified to species level through morphological identification using published keys[6]. When DNA barcoding was carried out on these specimens, the COI region could only identify *C. capitata*, *C. cosyra*, and *B. dorsalis* to species level. Sequence similarity between *C. rosa* and *C. quilicii* prevented differentiation based upon this gene region. DNA was successfully extracted from each specimen. DNA concentrations ranged from 25.4 to 320.0 ng/µl as determined by a Qubit dsDNA BR assay kit (Invitrogen). The DNA quality determined at the A260/A280 absorbance ratio on a NanoDrop 2000 spectrophotometer ranged between 1.9 and 2.12.

Gene selection, primer design, specificity, and sensitivity. De novo assembled contigs with high similarity to GenBank accessions XM_004526176.3 and XM_011215866.3 were targeted for species differentiation and primer design. One primer set was designed for each species with differing amplicon lengths for use in a multiplex PCR. Specificity tests performed on freshly extracted DNA from colony-reared insects showed

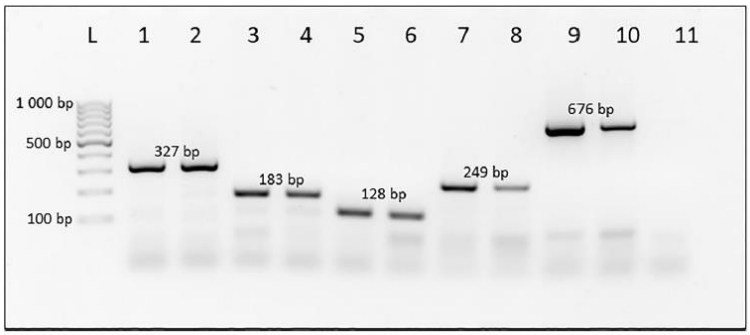


Figure 1. 2% agarose-TAE gel displaying the specificity of the multiplex PCR assay on freshly extracted DNA from colony-reared insects with species-specific amplicon size indicated. Lanes 1 & 2: *C. capitata*, Lanes 3 & 4: *C. cosyra*, Lanes 5 & 6: *C. quilicii*, Lanes 7 & 8: *C. rosa*, Lanes 9 & 10: *B. dorsalis*, Lane 11: No template control, Lane L: GeneRuler 100 bp DNA ladder (Thermo Scientific). The original gel is presented in Supplementary Fig. S1a.

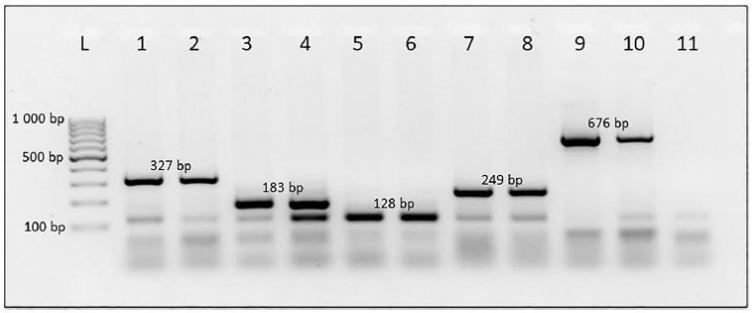


Figure 2. 2% Agarose-TAE gel displaying the specificity of multiplex primers in the case of duplex formation in freshly extracted colony-reared larval DNA with species-specific amplicon size indicated. Lane 4 demonstrates the expected *C. cosyra* amplicon at 183 bp with non-specific amplification at 128 bp leading to the formation of a duplex; the larger 183 bp amplicon should be used for identification. Lane 5 is a single 128 bp amplicon indicative of *C. quilicii*. Lanes 1 & 2: *C. capitata*, Lanes 3 & 4: *C. cosyra*, Lanes 5 & 6: *C. quilicii*, Lanes 7 & 8: *C. rosa*, Lanes 9 & 10: *B. dorsalis*, Lane 11: No template control, Lane L: GeneRuler 100 bp DNA ladder (Thermo Scientific). The original gel is presented in Supplementary Fig. S2a.

the presence of a single amplicon of the expected size for each species (Fig. 1). The results were consistent when tested on colony larvae, as shown in Fig. 2. A 2% agarose-TAE gel allowed for adequate separation of amplicons that were close in size for accurate species identification. Overall, the assay's detection limit was 10 ng and 4 ng when tested on colony larvae and colony adult DNA, respectively.

Assay validation on wild insects. The multiplex PCR assay was validated using freshly extracted DNA from wild, trap-collected specimens morphologically identified using available taxonomic keys[6]. It was noted that certain trap-collected specimens produced non-specific amplification of various sizes in addition to the expected identity amplicon. However, none of the non-specific amplicons interfered with the reliability or accuracy of the assay. An example of the efficacy of the multiplex PCR assay on trap-collected fruit flies can be found in Fig. 3.

Discussion

Five pairs of species-specific primers were designed, and a multiplex PCR was developed to identify fruit flies of economic importance in South Africa to species level. This assay generates a single amplicon of varying sizes for the different fruit fly species, *C. capitata* (327 bp), *C. cosyra* (183 bp), *C. quilicii* (128 bp), *C. rosa* (249 bp), and *B. dorsalis* (676 bp). These amplicons can be separated on a 2% agarose gel allowing for accurate differentiation without downstream analysis and sequencing. DNA concentrations of wild, trap-collected query specimens were not normalized during assay validation to demonstrate the efficacy of this assay for routine identification

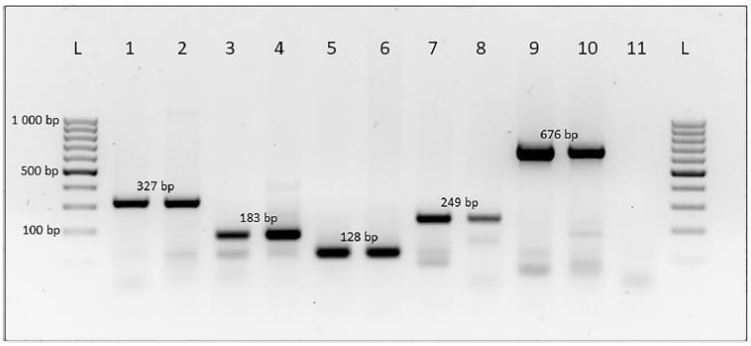


Figure 3. 2% agarose-TAE gel displaying the efficacy of the Multiplex PCR assay to identify wild, trap-collected specimens with examples of non-specific amplification. Species-specific amplicon sizes are indicated. Lanes 1 & 2: *C. capitata*, Lanes 3 & 4: *C. cosyra*, Lanes 5 & 6: *C. quilicii*, Lanes 7 & 8: *C. rosa*, Lanes 9 & 10: *B. dorsalis*, Lane 11: No template control, Lane L: GeneRuler 100 bp DNA ladder (Thermo Scientific). The original gel is presented in Supplementary Fig. S3.

where concentration normalization is not a priority, saving time when large numbers of specimens are being processed simultaneously.

In this study, all morphologically identified query specimens were correctly identified to species level using the multiplex PCR assay. This assay was developed for use as a differentiation tool to identify fruit fly pests of fresh fruit in South Africa that could potentially be present in export consignments and only validated to identify the five fruit fly pest species currently present in the country. The false-positive rate incurred when other fruit fly species are queried using this assay is unknown. Since other tephritid flies present in South Africa are not pests of commercial fruit primarily exported from South Africa, it is expected that only the five fly species investigated are likely to be intercepted on commercial fruit produced for export.

It was noted that in some specimens, a non-specific amplicon was present at the same size as the expected *C. quilicii* amplicon. The presence of this duplex is not to be confused with the *C. quilicii* amplicon, which will always yield a single 128 bp amplicon with no non-specific amplification. In cases where the duplex is present, the larger amplicon is to be used for species identification (Fig. 2). The species-specific primer pair designed for *C. quilicii* is located just outside of the opsin Rh4 coding domain in an intergenic region which has the potential to cross amplify in closely related species. Although the location of this primer set can lead to false-positive amplification in closely related species when used in conjunction with the other four species-specific primer sets designed in this study in a multiplex, accurate species differentiation of the five flies investigated can be achieved. The universal primer set CI-J2183 and TL2-N3014[34] can be used to amplify the COI gene, which can be sequenced and queried against the NCBI database for identity confirmation where applicable. It should be noted that CI-J2183 and TL2-N3014 can accurately identify *C. capitata*, *C. cosyra*, and *B. dorsalis* to species level. *Ceratitis capitata* and *Ceratitis caetrata* Munro (Diptera: Tephritidae) have previously been shown to share a high sequence similarity within the COI gene region, which may result in erroneous identification[35]. However, the distribution of *C. caetrata* is limited to Kenya, and the fly has never been reported in South Africa[36]. In closely related species and species complexes such as *C. rosa* and *C. quilicii,* this gene region does not allow accurate differentiation[3,23,29].

While the multiplex PCR assay designed in this study reliably performs its role, the downside of a multiplex is related to the use of PCR itself where the ability of the assay to reliably detect species present relies on the primer binding region to be conserved enough within the species so that any intraspecific variation present in the target region does not hinder amplification leading to false-negative results. This assay was validated with 15 wild insects per species collected from various sites across South Africa (Table S2) and no false-negative results were obtained. False-positive results can also occur where closely related species are highly similar to the target species leading to amplification. Fortunately, false-positive results are avoided in the multiplex assay described as species-specific amplicons are also size specific, so in cases of cross-amplification observed (Fig. 2) where a duplex is formed the larger amplicon is followed for accurate species identification. There is a high potential for false-positive results when other tephritid flies are queried against this assay, however, this assay is intended for use as a differentiation tool for identifying only the five fruit flies investigated.

The multiplex PCR detection assay developed in this study has application in identifying and monitoring agricultural pests of phytosanitary significance, both for pest management and surveillance practices. This relatively low cost and easy to perform assay uses only essential molecular laboratory equipment. It can be used in a standalone format or in conjunction with existing morphological identification techniques for improved accuracy in species identification. A significant advantage of this proposed method is that it allows for identification to species level without the need for downstream analysis. Reliable species identification can be achieved in under two and a half hours post DNA extraction, which significantly reduces the time required for existing molecular identification by DNA barcoding[37]. The increased turnaround time is a considerable advantage for inspection purposes in the implementation of a systems approach reducing the risk of fruit flies before fruit export, for

inspections of fruit consignments at ports of entry as well as for early detection of invasive fruit flies such as *B. dorsalis* which is currently absent in several areas in South Africa[38].

This assay was designed for fruit fly identification in the South African context to facilitate the identification of fruit flies of economic importance. However, given that many of these species also occur in other parts of Africa and are of economic importance in these regions, the assay may be of practical use in these regions as well. Further research will be required to determine the suitability of this assay for fruit fly identification in other African countries where other economically important tephritid flies occur, more specifically other members of the *Ceratitis* FARQ complex (*Ceratitis fasciventris* and *Ceratitis anonae*). Presently, the multiplex PCR assay developed in this study will provide a useful aid in decision-making regarding international trade and for monitoring and detection purposes.

Methods and materials

Sample collection, identification, and DNA extraction. Specimens used in this study were stored in 100% ethanol and kept at 4 °C until used. Colony insects and larvae came from established colonies held at Citrus Research International (CRI) in Mbombela, Mpumalanga, South Africa. Detailed information regarding the origin of the colonies is listed in Supplementary Table S1. The identities of the fruit fly species in the colonies (adult specimens from colonies refreshed in the period 2020–2021) were confirmed by Marc De Meyer, Royal Museum for Central Africa, on 21 February 2022. DNA was extracted from single insects following an adapted "salting out" protocol by Sunnucks and Hales[39], with TNES buffer (50 mM Tris, pH 7.5, 400 mM NaCl, 20 mM EDTA, 0.5% SDS) substituted for 180 µl ATL buffer (Qiagen) and incubation taking place overnight at 56 °C. Following the NaCl precipitation, 2 µl RNAse A was added to the supernatant and the second precipitation took place overnight at − 20 °C with isopropanol. DNA concentration and quality were quantified using a NanoDrop 2000 spectrophotometer and a Qubit dsDNA BR assay kit (Invitrogen).

Wild insects used for validation of the assay were collected from traps. Flies of the genus *Ceratitis* were trapped with McPhail type bucket traps baited with enriched ginger root oil (EGO lure) (Insect Science, Tzaneen, South Africa), and *B. dorsalis* flies were trapped with Chempac bucket traps baited with methyl eugenol (ME) (Invader lure, RiverBioscience, Gqeberha, South Africa). Total DNA was extracted from the whole body of the fruit fly following the destructive protocol of the DNeasy Blood and Tissue Kit (Qiagen).

The species of each adult colony specimen in this study was confirmed before the assay design using universal primer set CI-J2183 and TL2-N3014[34] for amplification and Sanger sequencing of the COI gene. The PCR was performed in a total volume of 25 µl containing 1 × Kapa Taq buffer A (KAPA Biosystems), 0.2 mM dNTP mix (Thermo Scientific), 0.4 µM of each primer (CI-J2183 and TL2-N3014), and 0.05 U KAPA Taq DNA Polymerase (KAPA Biosystems). The cycling conditions included an initial denaturation step at 94 °C for 5 min, followed by 35 cycles of 94 °C for 30 s, annealing at 50 °C for 30 s and elongation at 72 °C for 45 s. The final extension took place at 72 °C for 7 min.

High Throughput Sequencing and De novo assembly. DNA from two adult male specimens from the colony of each species were sent for high throughput sequencing at Macrogen (South Korea). Macrogen conducted library construction and high throughput sequencing of the colony insects on the Illumina NovaSeq 6000 platform (2 × 150 bp paired-end reads). Library preparation was performed using the TruSeq DNA PCR-Free Kit for the samples *C. rosa* 1, *C. quilicii* 1 & 2 and *C. cosyra* 2; and the TruSeq Nano DNA Kit for samples *C. capitata* 1 & 2, *C. rosa* 1, *C. cosyra* 1, and *B. dorsalis* 1 & 2, with input ranging from 0.565 to 2.998 µg of genomic DNA. De novo assembly was performed using CLC genomics workbench version 11.0.1 (Qiagen) and SPAdes[40] using default parameters as well as Velvet[41] with a hash length of 55.

Gene selection. Gene regions frequently used for differentiation of insect species were selected from literature and underwent preliminary bioinformatic analyses. A detailed list of these genes is available in Supplementary Table S3. The de novo assembled contigs were queried using BLAST + standalone (BLASTn algorithm) against a local copy of the NCBI GenBank nucleotide database. The gene regions of interest were then identified, and multiple sequence alignments were constructed to compare the genes between species using CLC genomics workbench version 11.0.1 (Qiagen). De novo assembled contigs with high similarity to GenBank accessions XM_004526176.3 and XM_011215866.3 (annotated as Opsin Rh3/Rh4) were targeted for species differentiation and primer design. This gene region showed the greatest potential for species identification due to the number of single nucleotide polymorphisms observed between species in the multiple sequence alignment. Literature suggests that the function of opsins within the order Diptera extends beyond visual processes influencing adaptation to new ecological niches and playing additional roles in host fruit detection, gustatory reception, and taste[42–46].

Primer design and multiplex PCR. A multiple sequence alignment of two reference sequences available on GenBank belonging to *C. capitata* and *B. dorsalis*, accessions XM_004526176.3 and XM_011215866.3 respectively, as well as de novo-assembled contigs high in similarity to these reference sequences for each species (GenBank accessions: ON505377–ON505386), was constructed. Five primer sets (IDT) were designed for differentiation of each species by amplicon size using Oligo Explorer 1.1.2 (Gene Link) (Table 1).

The PCR was performed in a total volume of 25 µl containing 1 × KAPA Taq buffer A (KAPA Biosystems), 0.2 mM dNTP mix (Thermo Scientific), 0.24 µM Opsin4_capitata_F & R, 0.32 µM Opsin4_cosyra_F & R, 0.32 µM Opsin4_quilicii_F & R, 0.64 µM Opsin4_rosa_F & R, 0.64 µM Opsin4_dorsalis_F & R and 0.05 U KAPA Taq DNA Polymerase (KAPA Biosystems). The cycling conditions included an initial denaturation step at 94 °C for 5 min, followed by 35 cycles of 94 °C for 30 s, annealing at 55 °C for 30 s and elongation at 72 °C for 35 s. The final extension took place at 72 °C for 7 min.

Primer pair	Sequence (5'–3')	Amplicon size (bp)
Opsin4_capitata_F	GCTAAAGCCATAACAATTCAG	327
Opsin4_capitata_R	CAGACTGTTCTTTTGGGC	
Opsin4_cosyra_F	GCTGTGACTTTGTTACAG	183
Opsin4_cosyra_R	GCATACTTGAATCTCAATCGAA	
Opsin4_quilicii_F	GCGTTCTGTTTTTAATCACTCA	128
Opsin4_quilicii_R	CATTTAATGTTTCAGAAGTGCT	
Opsin4_rosa_F	ATTGCTACAACTTTGTCGC	249
Opsin4_rosa_R	GCAGTAATACTGCGAATCATC	
Opsin4_dorsalis_F	TAGCACAATTATTTAGCGGG	676
Opsin4_dorsalis_R	ATTACCGTCAGCGATCAG	

Table 1. List of primers designed for accurate species identification in the multiplex PCR assay.

All visualizations of multiplex PCR amplicons in this study were separated on a 2% agarose TAE (2 M Tris, 1 M glacial acetic acid, 0.05 M Na_2EDTA, pH 8) gel stained with ethidium bromide.

To confirm that each specific primer pair generated the expected amplicons, each amplicon was bi-directionally Sanger sequenced with the relevant species-specific primer pair at the Central Analytical Facility of Stellenbosch University. A dilution series with a dilution factor of 5 was made with DNA extracted from both adult and larval colony specimens to determine the assay's sensitivity. The dilution series was quantified using the Qubit dsDNA BR assay kit (Invitrogen). The multiplex PCR assay was thereafter performed with the dilution series for adult insect DNA (20–0.0064 ng) and for larval DNA (50–0.0000256 ng) to determine the limit of detection.

Assay validation. The assay was validated using freshly extracted DNA from wild, trap-collected specimens of all five fruit flies, morphologically identified to species level using taxonomic keys. In total the assay was validated on 15 wild fruit flies of each species. 1 µl DNA was taken directly from the extract and used in the multiplex PCR without normalization for DNA concentration.

Data availability

The datasets generated and analyzed during this study are available in the NCBI GenBank repository, accession number: ON505377–ON505386.

Received: 21 April 2022; Accepted: 25 July 2022
Published online: 29 July 2022

References

1. White, I. M. & Elson-Harris, M. M. *Fruit flies of economic significance: their identification and bionomics*. (CAB international, 1992).
2. De Villiers, M., Manrakhan, A., Addison, P. & Hattingh, V. The distribution, relative abundance, and seasonal phenology of *Ceratitis capitata*, *Ceratitis rosa*, and *Ceratitis cosyra* (Diptera: Tephritidae) in South Africa. *Environ. Entomol.* **42**, 831–840 (2013).
3. De Meyer, M., Mwatawala, M., Copeland, R. S. & Virgilio, M. Description of new *Ceratitis* species (Diptera: Tephritidae) from Africa, or how morphological and DNA data are complementary in discovering unknown species and matching sexes. *Eur. J. Taxon.* **233**, 1–23 (2016).
4. Manrakhan, A., Venter, J. H. & Hattingh, V. The progressive invasion of *Bactrocera dorsalis* (Diptera: Tephritidae) in South Africa. *Biol. Invasions* **17**, 2803–2809 (2015).
5. De Meyer, M. Distribution patterns and host-plant relationships within the genus *Ceratitis* MacLeay (Diptera: Tephritidae) in Africa. *Cimbebasia* **17**, 219–228 (2001).
6. Virgilio, M., White, I. & De Meyer, M. A set of multi-entry identification keys to African frugivorous flies (Diptera, Tephritidae). *Zookeys* **428**, 97–108 (2014).
7. Malacrida, A. R. *et al.* Globalization and fruitfly invasion and expansion: the medfly paradigm. *Genetica* **131**, 1–9 (2007).
8. Qin, Y. *et al.* Population structure of a global agricultural invasive pest, *Bactrocera dorsalis* (Diptera: Tephritidae). *Evol. Appl.* **11**, 1990–2003 (2018).
9. Louzeiro, L. R. F., de Souza-Filho, M. F., Raga, A. & Gisloti, L. J. Incidence of frugivorous flies (Tephritidae and Lonchaeidae), fruit losses and the dispersal of flies through the transportation of fresh fruit. *J. Asia. Pac. Entomol.* **24**, 50–60 (2021).
10. Rasolofoarivao, H., Raveloson Ravaomanarivo, L. H. & Delatte, H. Host plant ranges of fruit flies (Diptera: Tephritidae) in Madagascar. *Bull. Entomol. Res.* **112**, 1–12 (2022).
11. PM 3/90 (1) Inspection of citrus fruits consignments. *EPPO Bull.* **50**, 383–400 (2020).
12. Whatson, M. Decision To Revise Import Requirements for the Importation of Fresh Citrus From South Africa Into the United States. *Fed. Regist.* **85**, (2020).
13. Bragard, C. *et al.* Pest categorisation of non-EU Tephritidae. *EFSA J.* **18**, (2020).
14. Balmès, V. & Mouttet, R. Development and validation of a simplified morphological identification key for larvae of tephritid species most commonly intercepted at import in Europe. *EPPO Bull.* **47**, 91–99 (2017).
15. Boykin, L. M., Armstrong, K. F., Kubatko, L. & De Barro, P. Species delimitation and global biosecurity. *Evol. Bioinforma.* **8**, EBO. S8532 (2012).
16. Virgilio, M. *et al.* An integrated diagnostic setup for the morphological and molecular identification of the Ceratitis FAR complex (C. anonae, C. fasciventris, C. rosa, C. quilicii, Diptera, Tephritidae). *Bull. Entomol. Res.* **109**, 376–382 (2019).
17. Tan, K. H., Wee, S.-L., Ono, H. & Nishida, R. Comparison of methyl eugenol metabolites, mitochondrial COI, and rDNA sequences of *Bactrocera philippinensis* (Diptera: Tephritidae) with those of three other major pest species within the *dorsalis* complex. *Appl. Entomol. Zool.* **48**, 275–282 (2013).

18. Jiang, F. *et al.* A high-throughput detection method for invasive fruit fly (Diptera: Tephritidae) species based on microfluidic dynamic array. *Mol. Ecol. Resour.* **16**, 1378–1388 (2016).
19. Jiang, F. *et al.* A conserved motif within cox 2 allows broad detection of economically important fruit flies (Diptera: Tephritidae). *Sci. Rep.* **8**, 2077 (2018).
20. Delatte, H., Virgilio, M., Simiand, C., Quilici, S. & De Meyer, M. Isolation and characterisation of sixteen microsatellite markers cross-amplifying in a complex of three African agricultural pests (Ceratitis rosa, C. anonae and C. fasciventris, Diptera: Tephritidae). *Conserv. Genet. Resour.* **5**, 31–34 (2013).
21. Virgilio, M., Delatte, H., Quilici, S., Backeljau, T. & De Meyer, M. Cryptic diversity and gene flow among three African agricultural pests: *Ceratitis rosa*, *Ceratitis fasciventris* and *Ceratitis anonae* (Diptera, Tephritidae). *Mol. Ecol.* **22**, 2526–2539 (2013).
22. Delatte, H. *et al.* Isolation and characterisation of sixteen microsatellite markers amplifying an African agricultural pest, *Ceratitis cosyra* (Walker) (Diptera: Tephritidae). *Conserv. Genet. Resour.* **6**, 9–11 (2014).
23. Zhang, Y. *et al.* Phylogenomic resolution of the *Ceratitis* FARQ complex (Diptera: Tephritidae). *Mol. Phylogenet. Evol.* **161**, 107160 (2021).
24. Virgilio, M., Backeljau, T., Barr, N. & De Meyer, M. Molecular evaluation of nominal species in the Ceratitis fasciventris, C. anonae, C. rosa complex (Diptera: Tephritidae). *Mol. Phylogenet. Evol.* **48**, 270–280 (2008).
25. Jiang, F., Jin, Q., Liang, L., Zhang, A. B. & Li, Z. H. Existence of species complex largely reduced barcoding success for invasive species of Tephritidae: a case study in *Bactrocera* spp. *Mol. Ecol. Resour.* **14**, 1114–1128 (2014).
26. Virgilio, M., Jordaens, K., Breman, F. C., Backeljau, T. & De Meyer, M. Identifying Insects with incomplete DNA barcode libraries, African fruit flies (Diptera: Tephritidae) as a test case. *PLoS ONE* **7**, e31581 (2012).
27. Barr, N. B., Islam, M. S., De Meyer, M. & McPheron, B. A. Molecular identification of *Ceratitis capitata* (Diptera: Tephritidae) using DNA sequences of the COI barcode region. *Ann. Entomol. Soc. Am.* **105**, 339–350 (2012).
28. Van Houdt, J. K. J., Bremean, F. C., Virgilio, M. & De Meyer, M. Recovering full DNA barcodes from natural history collections of Tephritid fruitflies (Tephritidae, Diptera) using mini barcodes. *Mol. Ecol. Resour.* **10**, 459–465 (2010).
29. Dhami, M. K., Gunawardana, D. N., Voice, D. & Kumarasinghe, L. A real-time PCR toolbox for accurate identification of invasive fruit fly species. *J. Appl. Entomol.* **140**, 536–552 (2016).
30. Barr, N. *et al.* Molecular diagnostics of economically important *Ceratitis* fruit fly species (Diptera: Tephritidae) in Africa using PCR and RFLP analyses. *Bull. Entomol. Res.* **96**, 505–521 (2006).
31. Barr, N. B., Garza, D., Ledezma, L. A. & Salinas, D. A. Using the rDNA internal transcribed spacer 1 to identify the invasive pest *Rhagoletis cerasi* (Diptera: Tephritidae) in North America. *J. Econ. Entomol.* **114**, 360–370 (2021).
32. Shariff, S. *et al.* Multiplex PCR in determination of Opiinae parasitoids of fruit flies, *Bactrocera* sp., infesting star fruit and guava. *J. Insect Sci.* **14**, (2014).
33. De Villiers, M. *et al.* The potential distribution of *Bactrocera dorsalis* : Considering phenology and irrigation patterns. *Bull. Entomol. Res.* **106**, 19–33 (2016).
34. Simon, C. *et al.* Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am.* **87**, 651–701 (1994).
35. Frey, J. E. *et al.* Developing diagnostic SNP panels for the identification of true fruit flies (Diptera: Tephritidae) within the limits of COI-based species delimitation. *BMC Evol. Biol.* **13**, 106 (2013).
36. de Meyer, M. Systematic revision of the subgenus Ceratitis MacLeay s.s. (Diptera, Tephritidae). *Zool. J. Linn. Soc.* **128**, 439–467 (2000).
37. Armstrong, K. F. & Ball, S. L. DNA barcodes for biosecurity: invasive species identification. *Philos. Trans. R Soc. B Biol. Sci.* **360**, 1813–1823 (2005).
38. Agricultural Pests Act, 1983 (Act No. 36 of 1983) Control measures: Amendment, in: Department of Agriculture, Forestry and Fisheries. *Government Gazette* 11–17 (2017).
39. Sunnucks, P. & Hales, D. F. Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). *Mol. Biol. Evol.* **13**, 510–524 (1996).
40. Nurk, S. *et al.* Assembling genomes and mini-metagenomes from highly chimeric reads. In *Research in computational molecular biology* (eds Deng, M. *et al.*) 158–170 (Springer, Berlin, Heidelberg, 2013).
41. Zerbino, D. R. Using the velvet de novo assembler for short-read sequencing technologies. *Curr. Protoc. Bioinforma.* **31**, (2010).
42. Feuda, R. *et al.* Phylogenomics of Opsin genes in Diptera reveals lineage-specific events and contrasting evolutionary dynamics in *Anopheles* and *Drosophila*. *Genome Biol. Evol.* **13**, (2021).
43. Sondhi, Y., Ellis, E. A., Bybee, S. M., Theobald, J. C. & Kawahara, A. Y. Light environment drives evolution of color vision genes in butterflies and moths. *Commun. Biol.* **4**, 177 (2021).
44. Leung, N. Y. *et al.* Functions of Opsins in *Drosophila* taste. *Curr. Biol.* **30**, 1367-1379.e6 (2020).
45. Feuda, R., Marlétaz, F., Bentley, M. A. & Holland, P. W. H. Conservation, duplication, and divergence of five Opsin genes in insect evolution. *Genome Biol. Evol.* **8**, 579–587 (2016).
46. Papanicolaou, A. *et al.* The whole genome sequence of the Mediterranean fruit fly, *Ceratitis capitata* (Wiedemann), reveals insights into the biology and adaptive evolution of a highly invasive pest species. *Genome Biol.* **17**, 192 (2016).

Acknowledgements
The authors thank Citrus Research International (CRI) for funding CRI project 1299 and providing the fruit fly specimens used in this study.

Author contributions
K.J.A. contributed to the study's design, primer design for the multiplex PCR assay, DNA extractions, HTS data analysis, optimization of the multiplex PCR assay, validation of the multiplex PCR assay, and drafting the manuscript. R.B. contributed to the study's design, primer design for the multiplex PCR assay, HTS data analysis, optimization of the multiplex PCR assay, and drafting the manuscript. A.M. contributed to the study's design, supplied all colony and trap-collected fruit flies, and drafting of the manuscript. H.J.M. contributed to the study's design and the manuscript's drafting. All authors read and approved the final manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-17382-x.

Correspondence and requests for materials should be addressed to H.J.M.