1hsh. 4j 0001

Morphological and genetic diversity of the new invasive species *Bactrocera invadens* (Diptera: Tephritidae) in Africa

Fathiya Mbarak Khamis

(BSc Biochemistry & Zoology, MSc Biochemistry)

Reg. No.: I84/15250/05

A thesis submitted in partial fulfillment of the Award of the Degree of Doctor of Philosophy (Molecular Biology) at the School of Pure and Applied Sciences of Kenyatta University

Khamis, Fathiya
Morphological and genetic diversity of

2011/359858

November 2009

KENYATTA UNIVERSITY LIBRARY

DECLARATION BY CANDIDATE

This thesis is my original work and has not been presented for a degree in any other University.

Signature:.....FMbarak..... Date:.....14/11/09....

Fathiya M. Khamis

DECLARATION BY SUPERVISORS

We confirm that the work reported in this thesis was carried out by the candidate under our supervision.

Signature:..... Date:.....14/11/09.....

Dr. Sunday Ekesi

African Fruit Fly Programme

International Centre of Insect Physiology and Ecology *(icipe)*

Signature:..... Date:.....24/11/09.....

Dr. Daniel Masiga

Molecular Biology and Biotechnology Depoartment

International Centre of Insect Physiology and Ecology *(icipe)*

Signature:..... Date:16th November 2009

Prof. Eucharia U. Kenya

Department of Biochemistry and Biotechnology

Kenyatta University

DEDICATION

To my family

ACKNOWLEDGEMENTS

First and foremost, I wish to thank God, the Almighty, for providing me with the opportunity, will and strength to undertake my studies. I recognize with great appreciation and gratitude the role played by my supervisors, Prof. Eucharia Kenya, Dr. Daniel Masiga and Dr. Sunday Ekesi in guiding me through this study. I am indebted to Dr Ekesi for facilitating my travel to Belgium and Italy to achieve the objectives of my study. I would also like to acknowledge the role played by Prof. Kenya in obtaining this fellowship at *icipe* and facilitating my university registration which was a requirement for the scholarship. I am grateful to Dr. Masiga for trusting in my capabilities. I appreciate the support of Prof. M. S. Rajab in the facilitation of my study. My heartfelt gratitude also goes to Dr. Samira Mohammed for her encouragement, advice and support during my most difficult time of the study. I also want to sincerely thank the Deputy Vice Chancellor (Administration) Kenyatta University, Prof. Geoffrey Muluvi for his support and encouragement. My appreciation also goes to Prof. A. Hassanali and Prof. M. Imbuga for their support and encouragement in pursuing my studies.

My sincere gratitude is to Dr. Marc De Meyer for his input to my proposal and invitation to his laboratory at the Royal Museum for Central Africa (RMCA), Belgium. I am also grateful for the African Biodiversity Information Centre (ABIC) for the fellowship to RMCA. I would like to acknowledge the Director of RMCA, the ABIC fellowship facilitators Ms Muirel van Nuffel and Ms Veronique Gilson, Ms Nathalie Erbout, all technical staff of the Entomology section of the RMCA, the Secretary of the department Ms Miriam, and the Information Technology department for their assistance.

My heartfelt thanks are also due to Prof. Anna Malacrida head of the Laboratory of Insect Evolutionary Molecular Biology of the University of Pavia (Italy) for the invitation to her laboratory and for her guidance and support. I would like to acknowledge Prof. Guilliano Gasperi, Dr. Ludvic Gomulski, Mr. Antonio, Dr. Nisrine Karam, Dr. Sabrina Bertin, Dr. Francesca Scolari and all other members of the Laboratory of Insect Evolutionary Molecular Biology of the University of Pavia (Italy). My sincere gratitude is also due to Prof. Rosalba Guglielmino for her assistance in the interpretation of the molecular analysis results.

I also extend my appreciation to the African Regional Postgraduate Programme (ARPIS) of *icipe* and the German Academic Exchange Services (DAAD) for the fellowship which enabled me to accomplish this academic achievement. The BMZ and the FORD foundation provided research funding through the African Fruit Fly Programme for which I am indebted. My appreciation also goes to the Chairman and colleagues in the Department of Biochemistry and Biotechnology, Kenyatta University.

My sincere thanks to my friends and colleagues at both Molecular Biology and Biotechnology Unit of *icipe* and AFFP, Mr. James Kabii, Ms Pamela Seda, Dr. Paul Mireji, Harisson Kibogo, Charity Wangechi, Nelly Ndungu, Catherine Gambi, Peter Kuria, Joel Bargul, Celestine Dweni, Pauline Asami and the other 'Girl children', Steven Ger, Dr. Eric Koum, Dr. Martin Rono, Aman Bonaventure, Eunice Machuka, Mark Kimondo, Peris Amwayi, Beatrice Gikaria, Peris Machera, Joseph Gitau, Peterson Nderitu, Wanyoni Martin, Emmanuel Mlato, John Kiilu, Vincent, Tanga Mbi, Felix

KENYATTA UNIVERSITY LIBRARY

Appiah, Piera Muthoni and my ARPPIS colleagues. My Special thanks also go to the staff of *icipe* Capacity Building, Lillian Igweta, Lisa Omondi and Margaret Ochanda. Edda Wasike, Wellington Ambaka and Joash Olago were extremely helpful in literature search at the library. Thanks are also due to my friends Yusuf Abdillah and Fikira Kimbokota. I would also like to thank all the people who sent us insects from different localities, this was greatly appreciated. Thanks are due to Dr. M. Billah for identification of the insect samples for this study and assistance of data analysis of the morphometric results.

Last but not least, I wish to express my heartfelt gratitude to my family for the encouragement and support throughout my study.

TABLE OF CONTENTS **PAGES**

KENYATTA UNIVERSITY LIBRARY

KENYATTA UNIVERSITY LIBRARY

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Species of the genus *Bactrocera* in the family Tephritidae, or 'true fruit flies', are among the most important pests of fruits and vegetables. Because of the broad larval host range and cosmopolitan distribution, pest fruit flies are highly invasive, with adults often exhibiting a strong tendency for dispersal and larvae being readily transported to new areas via fruit movement. Recently, a pest species was recorded for the first time in Kenya and has subsequently been found in countries across tropical Africa. The insect was described as *Bactrocera invadens*, due to its rapid invasion of the African continent. Taxonomic description of the *B. invadens* depicted different thoracic colourations that are morphotypes of the same pest. As invasive species spread around the globe, it is becoming increasingly evident that a detailed knowledge of the biology, genetic structure and geographical variability, of a given species is a prerequisite to planning strategies for quarantine, control or eradication. In this study, the morphometry, population structure and the genetic variability of different populations of the *B. invadens* distributed across the actual species range of tropical Africa and a sample from the presumed aboriginal home of Sri Lanka was investigated. Morphometry using wing veins and tibia lengths were used to separate the morphotypes of the *B. invadens* and a comparison to other closely related *Bactrocera* species. Fourteen distances between 15 selected landmarks on the wing were computed to characterize the shape and size of the wings for differentiation. *Bactrocera invadens* from all the localities analysed could not be separated by both the Principal component and canonical variate analyses. These two parameters separated the *B. invadens* from *B. correcta*, *B. cucurbitae*, *B. oleae* and *B. zonata* and it clustered together with *B. dorsalis* and *B. kandiensis*. Eleven polymorphic microsatellite loci were isolated and characterized and used for population genetic studies. The polymorphism of these loci was tested in individual flies from two natural populations (Sri Lanka and Democratic Republic of Congo). Allele number per locus ranged from three to 15 and eight loci displayed a polymorphic information content greater than 0.5. These markers were used to address three major issues on 13 geographic populations of the *B. invadens*, namely: (i) The historical origin of the *B. invadens*, (ii) the origin of the invasion of the *B. invadens* in Africa, and (iii) the extent of the establishment of the *B. invadens* populations in Africa and the genetic variability of these populations. Few years after its discovery in East Africa and its identification as a member of *Bactrocera dorsalis* complex, *B. invadens* samples from all equatorial Africa showed a high level of genetic diversity associated with an evident absence of geographic structure. These features are indicative of processes of rapid population growth and expansion with possible multiple introductions. The association of these two aspects reveals that *B. invadens* was not an indigenous African species that remained undetected for long period of time. Instead this species is a recent and an aggressive invader. Data from the timing of historical records, that indicated this insect as a new entrant in a region, are in this case concordant with the chronology of the spread as documented in this study. DNA barcoding using the CO1 gene was used to investigate the identity and species integrity of the *B. invadens* by comparing it with other *Bactrocera* species such as *B. correcta, B. cucurbitae*, *B. dorsalis*, *B. kandiensis*, *B. oleae, B. paraverbascifoliae* and *B. zonata*.

CHAPTER ONE

INTRODUCTION

1.1 Background

Sub-Saharan Africa is characterized by diverse climatic conditions that are favourable for horticultural production. Cultivation of fruits and vegetables especially serves as an important source of income for many small holder farmers who produce for both domestic and export markets, also in the improvement of diet by providing essential micronutrients and vitamins as well as creating job opportunities. However, horticultural production in Africa is generally believed to be far below its potential. This is largely attributed to several biotic and abiotic constraints. Among the abiotic constraints are poor availability of planting materials, and infrastructure, and high cost of inputs (Wessel, 1997). Biotic factors include among others, heavy attack by several insect pests among which infestation by fruit flies (Diptera: Tephritidae) rank very high (Lux *et al.,* 2003b).

Fruit flies are picture-winged flies of variable size with worldwide distribution. This family includes more than 4000 species in nearly 500 genera that are distributed within the temperate, subtropical and tropical areas of the world (White and Elson-Harris 1992; Norrbom *et al.*, 1999). Although approximately 1400 species are known to develop in fleshy fruits, larval development can take place in other parts of the plant including flowers, seeds and stems (De Meyer, 2001; Copeland *et al.*, 2002). Nearly 250 of these species are capable of achieving pest status by feeding on plants of economic importance (White and Elson-Harris 1992; Drew and Hancock 1994). Some Tephritid fruit flies are highly polyphagous and are economically significant pests of agriculture (Prokopy and

Roitberg; 1984), while many are minor or potential pests (White and Elson-Harris 1992; Thompson, 1998). Direct losses to fruits and vegetables are caused by female flies that lay eggs under the skin of fruits and vegetables and developing larvae feed on the flesh. This results in rotting of the fruits. In the tropics, the problem is aggravated by the prevailing warm weather, which is conducive to asynchronous fruiting patterns, resulting in overlapping generations and the potential of infestation all the year around (Purcell, 1998). Assessment by Lux *et al.* (2003b) revealed that of the 1.9 million tonnes of mangoes produced in Africa annually, about 40% is lost due to fruit flies. In addition to this direct loss, major indirect losses result from quarantine restrictions that are imposed by importing countries to prevent entry and establishment of unwanted fruit flies. This leads to loss of markets to farmers.

Africa is the aboriginal home of several well known fruit flies, such as the Mediterranean fruit fly, *Ceratitis capitata* (Wiedemann), the Natal fruit fly, *C. rosa* Karsch, *C. anonae* Graham, the Mango fruit fly, *C. cosyra* (Walker) and several other species in the genera *Ceratitis* MacLeay, *Trirhithrum* Bezzi and *Dacus* Fabricius (Cogan & Munro, 1980; Munro, 1984; White & Elson-Harris, 1992; De Meyer, 1996, 1998, 2000). Due to poor quarantine a service, Africa is also particularly vulnerable to invasive fruit fly species. In 2003 an unusual fruit fly species of the genus *Bactrocera* previously thought to be *Bactrocera dorsalis* Hendel was detected in coastal Kenya (Lux *et al.*, 2003a) and then Tanzania (Mwatawala *et al.*, 2004) and was recently described by Drew *et al.* (2005) as *Bactrocera invadens*. This species name arose from its rapid invasion of the African continent. In addition to Kenya, the pest has now been recorded from Benin, Cameroon,

KENYATTA UNIVERSITY LIBRARY

Comoros island, DR Congo, Equatorial Guinea, Ghana, Guinea, Mali, Nigeria, Senegal, Sudan, Tanzania, Togo, Uganda and the Comoros Island (Drew *et al.*, 2005; French, 2005; S. Ekesi, unpublished data: R. Hanna, unpublished data). It is also known from Sri Lanka (Drew *et al.*, 2005). *B. invadens* is believed to have invaded Africa from the Indian subcontinent and was recognized as a different species from Sri Lanka after it was first reported from Africa, where it has become a significant pest of quarantine and economic importance (Mwatawala *et al.*, 2004; Ekesi *et al.*, 2006). Taxonomic description by Drew *et al.* (2005) depicted different thoracic colourations of *B. invadens*. These different colour forms are morphotypes of the same pest.

In the majority of tephritid fruit fly species, diagnostic morphological characters for the identification of adult flies are now available (White & Elson-Harris, 1992). However, morphological tools present some limitations. Firstly, because of high homoplasy in most morphological characters and the existence of cryptic species groups across the family, the classification of Tephritids to the species level based on adult morphology alone is difficult (Armstrong *et al.*, 1997; De Meyer, 1999; McPheron, 2000). Also importantly, larvae cannot be reliably identified on morphological characters and yet most interceptions are of larval stages (Armstrong *et al.*, 1997). This problem is further compounded by the fact that morphological variation can be affected by environmental factors and genotypic environmental interactions (Bubliy *et al.*, 2001; Bubliy & Loeschcke, 2002). These limitations have led several taxonomists and quarantine officials alike to seek viable alternative ways of fruit fly identification including the use of molecular markers (Sonvico *et al.*, 1996; Armstrong *et al.*, 1997; Morrow *et al.*, 2000).

KENYATTA UNIVERSITY LIBRARY

1.2 Problem Statement and Justification

One of the major factors limiting increased production of fruits and vegetables is infestation by fruit flies. A complex of native fruit fly species including *Ceratitis cosyra*, *C. fasciventris*, *C. rosa*, *C. anonae* and *C. capitata* are responsible for the damage. For example, assessments made by partners of the *icipe*-led African Fruit Fly Initiative revealed that, out of 1.9 million tones of mangoes produced in Africa annually, about 40-70% is wasted due to fruit flies. The infestation rate varies among countries, seasons and variety. Quarantine restrictions on fruit flies restrict export to large lucrative markets in Europe, the Middle East, Japan and USA, where these insects are quarantine pests. With the invasion of *B. invadens*, the fruit fly problem has been compounded.

Preliminary results from host range studies carried out at *icipe* on the *B. invadens* show that *B. invadens* is polyphagous, attacking both cultivated and wild hosts including mango *Mangifera indica* L. (Anacardiaceae), lemon orange *Citrus limon* (Burman f.) (Rutaceae), tomato *Lycopersicon esculentum* Miller (Solanaceae), banana *Musa* sp., (Musaceae), guava *Psidium guajava* L. (Myrtaceae), marula *Sclerocarya birrea* H. (Anacardiaceae), custard apple *Annona muricata* L. (Annonaceae), avocado *Persea americana* M. (Lauraceae) and Indian almond *Terminalia catappa* L. (Combretaceae). It is already feared that the geographical expansion and infestation could be more severe if no control action and efficient quarantine systems are put in place. Identification and taxonomic placement of the pest will enable proper machinery of management be operational to reduce the risk of spread of the pest and the host range.

Bactrocera invadens has been observed to exhibit different thoracic colourations that are morpho-types of the same pest. These colour forms have been found in all localities and countries where it was collected. Besides the co-existence of the morphotypes, the problem of separation of *B. invadens* is further exacerbated by the closely related nature and resemblance of the insect to *B. dorsalis s.s*, *B. kandiensis*, and many other Asian *Bactrocera* species. The morphometric and molecular studies are crucial for a clear understanding of the behaviour across geographical range and interaction of local populations of the pests with natural enemies. Research data on the genetic structure of fruit fly populations: their intrinsic variability, gene flow and genetic inter-relationships should provide developers of control tools/strategies with the needful template for identifying their native habitat(s) – the potential source areas of biological control agents, tracing the route(s) of their colonization from the source area(s), evaluating the differentiation in the various populations, gaining insight into their dispersive potentials and mating compatibilities, and understanding the evolution of host plant use. Moreover, genetic information can enable prediction of possible responses of natural populations to control measures.

The study proposes to clarify the complexity of the pest population structure, starting with separation of the different colour forms and determining any morphometric or genetic differences, and how they relate to other invasive or potentially invasive *Bactrocera* species such as *B. caryea, B. correcta*, *B. cucurbitae*, *B. dorsalis s.s, B. kadiensis, B. oleae* and *B. zonata*. This will help ascertain the species limit, facilitate their reliable separation and enhance the future development of targeted management options.

1.3 Objectives of the study

1.3.1 General objective

The overall objective is to describe the population structure of *Bactrocera invadens* across its known distribution range in order to contribute to its control.

1.3.2 Specific objectives:

1. To investigate the peculiarity status of the morphotypes of *B. invadens* in Africa and its presumed aboriginal home (Sri Lanka) and establish the species limit through morphometric studies with comparisons with six closely related *Bactrocera* species (i.e *B. correcta, B. cucurbitae*, *B. dorsalis s.s, B. kandiensis, B. oleae* and *B. zonata*).
2. To conduct molecular studies on the different morphotypes of *B. invadens* to ascertain their genetic status.
3. To establish the inter-country and within region and country variability of the most dominant morphotype of the *B. invadens* and assess its colonization and invasion history.
4. To investigate the population genetic structure of *B. invadens* and compare it with closely related *Bactrocera* species such as *B. correcta, B. cucurbitae*, *B. dorsalis s.s*, *B. kandiensis*, *B. oleae B. paraverbascifoliae*, and *B. zonata.*

CHAPTER TWO

LITERATURE REVIEW

2.1 Life history

Female adult fruit flies lay their eggs beneath the skin of suitable hosts, especially in physiologically mature, ripening or ripe fruits, depending on the fruit fly species and the host plant attacked. The eggs are laid singly or in cluster. Some species such as *C. capitata*, and several *Anastrepha* and *Rhagoletis* species, have been shown to use oviposition deterrent pheromone as signal to con-specifics that the fruit has already been attacked (Averill and Prokopy, 1989). Eggs hatch in about 2-12 days depending on the species. The hatching larvae shed their skins twice as they feed and grow, and third instars larva emerges from the fruit and drops to the ground. The larvae of most fruit feeders can "pop" on the ground finding suitable sites for pupariation. The formation of visible structure that follows the last larval instars in Tephritidae is often erroneously called pupation but it is actually pupariation as true pupation takes place unseen within the puparium (White & Elson-Harris, 1992). At the completion of the third instars, the larval skin hardens to form a puparium with inactive fourth-instars larvae inside (Christenson and Foote, 1960). Eventually the larva within the puparium sheds its skin, forms a pupa, from which the adult fly later emerge. The pupal stage takes between 6 days (in *C. capitata*) to 30 days (in *B. dorsalis*). The emerging adults crawl upward through the soil usually at an angle using cracks and crevices that egress to the surface, especially when the soil is hard and compact (Christenson and Foote, 1960). Adult life span ranges from one month to as long as eleven months in *A. ludens* (Loew) (White & Elson-Harris, 1992). Newly emerged adults require a carbohydrate energy source and water for survival. In addition, the female search for a protein source for

egg maturation. After mating and a pre-oviposition period, which varies with species, the female starts laying eggs, and a new cycle begins. The duration of the different stages also varies with host plant and climatic conditions. Tropical species in the genera *Ceratitis*, *Bactrocera*, *Dacus* and *Anastrepha* are multivoltine i.e. they have several generations per year.

2.2 Fruit fly distribution, invasions and host plant relationships

The family Tephritidae is represented in all regions of the world, except the Antarctica. The major pest genera each have a limited natural distribution. For example, *Ceratitis* species attacks a wide range of fruits and are native to tropical Africa. *C. cosyra* is broadly distributed across Eastern, Central and Western Africa, and also in parts of Southern Africa and it attacks mangoes, marula, custard apple and guava (White & Elson-Harris, 1992; De Meyer, 2001). *C. rosa* is broadly distributed throughout Eastern and Southern Africa while *C. fasciventris* is scattered throughout Central and Western Africa (De Meyer, 2001). Both pests are highly polyphagous and attack other cultivated fruits apart from mangoes. *C. anonae* is broadly distributed in Western and Central Africa and its host range includes mango, coffee, tropical almond, common guava and strawberry guava (White & Elson-Harris, 1992; De Meyer, 2001). *C. capitata* is established in most regions of the world except Asia, whilst several detections and outbreaks in North America have occurred but have been successfully eradicated. *Bactrocera* species (formerly included in *Dacus*) are native to tropical Asia, Australia and the South Pacific regions, with a few species found in Africa and warm temperate areas of Europe and Asia America. One section of the genus, *Bactrocera (Zeugodacus)*, is most exclusively associated with

flowers and fruits of Cucurbitaceae, while the rest of the genus is associated with a wide range of fruits predominantly of tropical west region. Some *Bactrocera* species *B. dorsalis* complex have become established in Hawaii, California, French Guiana, Indian Ocean Island, Egypt and Suriname probably due to increased fruit trade and movement of contraband fruits. *Dacus ciliatus* (Loew) has become established in the Indian subcontinent and the Indian Ocean Islands. *Dacus* species are almost associated with flowers and fruits of Cucurbitaceae, or with the pods of Asclepiadaceae, and most species are found in Africa. *Anastrepha spp.* attacks a wide range of fruits in South and Central America and the West Indies, with a few species occurring in the extreme south of the USA. No *Anastrepha* species have become established outside those areas (White & Elson-Harris, 1992).

2.3 Control strategies of fruit flies

Different control strategies are used to combat the fruit fly menace. These include physical, cultural, behavioral, genetic, chemical and biological (mainly parasitoids) methods. However, the greatest reliance has been on chemical control. Chemical insecticides (whether blanket spray or in bait form) may have negative effects on the environment and on non-target organisms, and can be hazardous to the applicators. On the other hand, developing food attractants for African fruit flies could be a difficult task particularly in their habitat where there are abundant natural food source.

A more ecologically acceptable control measure is the sterile insect technique (SIT), but this approach is complex and very expensive (Bateman, 1972). SIT may not work as a sole

control strategy, particularly when the population density of the fruit flies is high (Knipling, 1992), and perhaps more importantly, when several species co-exist. Moreover, when considering the cosmetic value of fruits, the use of the SIT may not be compatible with grower requirements, because sterile females will continue to oviposit and damage fruits, even if the eggs are not viable (Sivinski *et al.*, 1996). Furthermore, SIT for control of *C. cosyra*, *C. rosa*, *C. fasciventris*, and *C. anonae* is currently not possible because no appropriate methods for mass production of these species have been developed. Additionally, horticultural production in African countries is not sufficiently developed to justify such expensive technology.

Biological control is considered environmentally acceptable and economically feasible especially classical biological control, which does not involve additional cost, once natural enemies have been established. Moreover, the use of parasitoids in combination with other compatible methods, i.e. sterile insect techniques, represents a sound and viable option in fruit fly management (Knipling, 1992). The use of parasitoids in most fruit fly biological control programs follows a classical approach (Clausen, 1978; Wharton, 1989), whereby the parasitoids that attack fruit flies in their aboriginal home are imported and released into the invaded areas. Different parasitoid guilds, i.e. egg, larval and pupal parasitoids, are normally included during importation programs, as well as a high diversity including species from the following families: Braconidae, Chalcididae, Diapriidae, Eulophidae, Eupelmidae, Eurytomidae, Eucoildae, Ichneumonidae, and Pteromalidae (Wharton, *et al.*, 1981; Hoffmeister, 1992). However,

KENYATTA UNIVERSITY LIBRARY

the Braconidae, particularly the subfamily Opiinae contains the largest diversity of parasitoids attacking tephritid fruit flies (Clausen, 1978).

Biological control agents other than parasitoids can be used for control. Microbial agents such as fungi of genera *Penicillium*, *Serratia* and *Mucor* have been reported to cause considerable larval and pupal mortality to *B. dorsalis* (Newell & Haramoto, 1968). A study at *icipe*, Kenya, by Ekesi *et al.* (2002) also revealed that several isolates of *Metarhizium anisopliae* (Metsch.) Sorok and *Beauveria bassiana* (Balsamo) Vuillemin were found to be effective against adults, larvae, and puparia of a number of African fruit fly species (*C. capitata, C. rosa, C. cosyra, C. fasciventris* and *C. anonae*).

2.4 *Bactrocera* species in Africa

The genus *Bactrocera* has 350 to 376 species with some undescribed species remaining in collections (Lawson *et al.,* 2003). *Bactrocera* species (formerly included in *Dacus*) are native to tropical Asia, Australia and the South Pacific regions, with a few species found in Africa and warm temperate areas of Europe and America (White & Elson-Harris, 1992). With adult traits that include high mobility and dispersive powers, high reproductive rates, and extreme polyphagy, *Bactrocera* species are well-documented invaders and rank high on quarantine lists world-wide (Clarke *et al.*, 2005). Asian fruit fly pests from this genus are regarded as one of the most destructive insects of fruits and vegetables worldwide (White & Elson-Harris, 1992). Losses are due to direct feeding damage and loss of export market opportunities through quarantine restrictions imposed by importing countries to avoid entry and establishment of fruit fly pests. The *Bactrocera dorsalis*

KENYATTA UNIVERSITY LIBRARY

Hendel complex of tropical fruit flies contains 75 described species, most endemic to Asia (Weems *et al.*, 1999; Clarke *et al.*, 2005). The group is arguably one of the most important pest species complexes in world agriculture (Clarke *et al.*, 2005; Drew *et al.*, 2005).

Previous *Bactrocera* species found in Kenya included *B. cucurbitae*, *B. bigulatta* (Bezzi), *B. munroi* White and *B.oleae* (Copeland *et al.*, 2004). *Bactrocera cucurbitae* was the most frequently encountered *Bactrocera* species in most eastern African countries and has recently been reported in Gambia and Ghana in West Africa (Norrbom, 2002; Billah, unpublished data). *B. dorsalis* (Hendel) was briefly established in the Indian Ocean island of Mauritius was but declared eradicated in 1996 (Seewooruthun *et al.*, 1998), while *B. zonata* (Saunders) was detected in Egypt, Mauritius and La Réunion in 2000 (Norrbom, 2002). *B. invadens* was discovered in Africa in 2003 and is now widely distributed in sub-Saharan Africa (Drew *et al.*, 2005; Mwatawala *et al.*, 2004; Lux *et al.*, 2003).

Eradication of *Bactrocera* spp is technically possible as demonstrated in the United States (U.S) and Australia though at a tremendous cost (Weems *et al.*, 1999). Despite the resources and strong quarantine facilities, interceptions are common at ports of entry into the U.S suggesting potential for outbreak. Weems *et al.*, (1999) also points out that eradication of *B. cucurbitae* was achieved in Japan over a period of 18 years in an extensive program involving insecticides, male attractants and sterile insect technique. Such actions however require rigorous operational standards and huge investments. However, the scale of the already known distribution of *B. invadens* in Africa, the

fragmented structure of farms, poor quarantine services, vast reservoir and lack of state or donor resources makes eradication a difficult if not an impossible task (Ekesi *et al.*, 2006).

2.5 Taxonomic analysis

2.5.1 Morphometrics

Morphometrics is the quantitative description or the analysis of measurements (or ratios of measurements) of form, shape and structure (morphology) in an organism. The morphological characters used in morphometric analysis include tibia length, tibia-ovipositor ratio, wing vein lengths and many other features which can differentiate populations. These measurements produce continuous variables (Sokal & Rohlf, 1995) and are usually taken from well-defined landmarks, which must be repeatable, homologous, and carefully described (Bookstein *et al.*, 1985). Geometric morphometrics using microscopy and camera Lucida for drawing features like the wing veins at defined landmarks is a novel technique.

Morphometric methods are frequently used for identification of populations within species and it is often necessary to use more than one character to discriminate between groups. According to Richtsmeier *et al.* (2002), the first approach involves the application of quantitative methods to discover new information within the data and then looking for patterns within the data that suggest any underlying biological processes or effects. To analyse such large datasets, computer programmes of multivariate methods are needed (Sneath & Sokal, 1973), and some of the commonest methods include (among others) principal components and canonical variates analyses.

Principal component analysis (PCA) and Canonical variates analysis (CVA) are techniques used in multivariate statistics and systematic to examine the information in complex multidimensional data sets (Woolley *et al.*, 1994). Principal component analysis also examines relationships among several quantitative variables and has been viewed as an attempt to uncover approximate linear dependencies among variables (SAS Institute Inc., 2001).

Principal components (PCs) have a variety of properties that make them appropriate or useful for morphometry. The components are sorted by descending order of the eigenvalues, which are equal to the variances of the components (Rao, 1964; Kshirsager, 1972). The first principal component (PC 1) has the largest variance of any unit-length linear combination of the observed variables; the second principal component (PC 2) has the next largest variance of what variance is left in the data, and so on to the last principal component (Marcus, 1990; Rolf & Bookstein, 1990), which is the most invariant combination of the data, i.e has the least variance. Each principal component is a linear combination of the original variables, with coefficients equal to the eigenvectors of the correlation or covariance matrix (SAS Institute Inc., 2001).

Canonical variates analysis represents a new coordinate system but, in this case, the new dimensions are constructed so as to maximally discriminate between pre- assigned groups. It is the multigroup extension of linear discriminant analysis in which a single axis is constructed to maximally discriminate between two groups (Rohlf & Bookstein, 1990; SAS Institute Inc., 2001). As in linear discriminant analysis, the first canonical

variate (CV 1) is constructed to maximize between group covariance and minimize within group covariance (Woolley *et al.*, 1994). The second canonical variate (CV 2) provides maximal discrimination, subject to the constraint that it is orthogonal to the first, and so on. The distance between any two centroids in canonical variates space is known as the Mahalanobis distance, and usually presented in the squared form (D^2). According to Woolley *et al.* (1994), it provides the best multivariate measure of the relative distance between groups taking into account the variation within each one of them.

A study on the fruit fly parasitoids by Kimani-Njogu *et al.* (2001) used morphometrics and hybridization techniques to establish the identity of *Psyttalia* populations from Rurima and Ruiru attacking *C. capitata* on coffee in Kenya. Morphometric analysis has also been used to identify and characterise populations of African *Psyttalia* to facilitate their reliable use as biological control agents of fruit infesting flies (Diptera: Tephritidae) in Africa by Billah *et al.* (2005).

2.6 Molecular methods in fruit flies population genetics

Uncertainty in species limits based on the traditionally used adult morphological features, together with overlapping host and geographical ranges, significantly impacts quarantine, pest management, and general biological studies (Clarke *et al.*, 2005). Accurate identification is essential for fruit fly species found in fruits destined for export, for distinguishing exotic from native fauna, and for providing crucial data on risk and invasion pathways (Clarke *et al.*, 2005). These limitations have forced quarantine officials and taxonomists to seek alternative ways to fruit fly identification (Sivinski *et al.*, 1996;

Armstrong *et al.*, 1997; Morrow *et al.*, 2000), classification (Han & McPheron, 1997; 1999) and to understanding population structure (Malacrida *et al.*, 1998) through the use of molecular markers. Molecular markers can overcome limitations attributed to morphological characters, since they are phenotypically neutral and hence life-stage independent and impervious to environmental conditions, including the possible physical damage to the specimens (Baliraine *et al.*, 2002). Several molecular markers are available for studying the variation within and between species of insects and assessing the genetic structure of insect populations. These have been variously used to delineate species, determine insect phylogenetic relationships (Caterino and Vogler, 2002) and elucidate intraspecific variation (Omondi *et al.*, 2003; Berry *et al.*, 2004; Billah, 2004). Molecular markers for such exploratory studies should be those that either requires no prior sequence information for the species (e.g. RAPD-PCR and AFLP) or those for which probes may be obtained from other closely related studied species e.g. mitochondrial DNA (mtDNA) markers.

2.6.1 PCR-RFLP analysis of Mitochondrial DNA

The development of molecular methods for identifying fruit fly species aim to overcome limitations of identification based on morphology alone. Among loci that have been used is mitochondrial DNA. In animals, the mitochondrial genome is generally circular (15 – 17 kb), maternally inherited and has a relatively simple genetic structure and rapid rate of sequence change (Thao *et al.*, 2004). The DNA molecule is made up of 37 coding genes encompassing 22 transfer RNAs, 2 ribosomal RNAs and 13 messenger RNAs, with a general lack of introns, large families of repetitive DNA, pseudogenes and large spacer

sequences (Loxdale and Lushai, 1998). The gene order of major regions is: ND2, COI, COII, ATP, C3, N3, ND5, ND4, ND6, Cytb, ND1, 16S, 12S and the A+T regions (corresponding to the sequences of *Drosophila yakuba* Burla (Diptera: Drosophilidae) (Clary and Wolstenholme, 1985).

The mtDNA is probably the most widely used source of DNA for molecular investigations and a good starting point for preliminary investigations and has been used for taxonomic and population genetic studies of insects (Loxdale and Lushai, 1998). Apart from versatility of its applications, the mtDNA markers can differentiate to the level of race forms. This is partly because mtDNA mutate approximately 20 times faster than nuclear DNA (Brown *et al.,* 1979). Restriction sites associated with these specific gene regions are highly conserved and are used for systematic studies of species and higher taxa. Furthermore, numerous molecules of circular mtDNA are found within a single cell and the chance of finding intact portions of the molecule from a degraded specimen is much higher than for a single-copy nuclear gene.

Recent development using nucleotide sequence data of certain genes of interest has provided a new approach to molecular systematics and a possibility for new classification. Known conserved regions or genes are Polymerase Chain Reaction (PCR) amplified using conserved oligonucleotide primers and sufficient DNA copies are obtained from a single fly specimen for direct sequencing, thereby replacing the time-consuming gene-cloning process. Sequence analysis of these regions may be followed by restriction digestion analyses on an agarose or polyacrylamide gel to visually reveal intra-specific variation and

polymorphic microsatellites. Sequence analysis of reference samples may also be used to improve selection of appropriate endonucleases that recognize diagnostic sites for species identification though selection can be performed by trial and error (Quinteiro *et al.*, 1998; Cervera *et al.*, 2000). The technique is known as restriction fragment length polymorphism analysis of PCR amplified DNA (PCR-RFLP). This method is quick, sensitive, life-stage independent and selectively neutral, and may be useful for generating species diagnostic markers. Sheppard *et al.* (1992) used RFLP of the mtDNA based on two enzymes to differentiate Medfly populations from different localities.

2.6.2 Microsatellites (Simple Sequence Repeats (SSR))

Microsatellite sequences are short core (1-6 bp) tandem repeats up to 500 bp long scattered throughout the genome (Weber and May, 1989). Uniquely, many are closely associated with conserved loci containing coding regions (Loxdale and Lushai, 1998). These markers are codominant and hypervariable, revealing many alleles per locus, and inherited in Mendelian fashion hence able to refine detectable genetic variability. Once primers have been developed, they may be used in closely related taxa. It has been recently proposed that selectively neutral genetic markers, such as microsatellite loci, are the most suitable for determining the historical distinctiveness of populations, and hence, the designation of species (Hedrick *et al.*, 2001). Microsatellite can be amplified even from highly degraded DNA sources, and are easily scorable (Bruford & Wayne, 1993). Moreover, being highly polymorphic, they can be utilized in the analysis of the fly population structures across different geographical areas, and in tracing of the origins of colonist populations (Bonnizoni *et al.*, 2000, 2001; Meixner *et al.*, 2002). The origin of

such polymorphism is still under debate though it appears most likely to be due to slippage events during DNA replication (Schlötterer & Tautz, 1992). Despite the fact that the mechanism of microsatellite evolution is still unclear, SSRs have been widely employed in many fields soon after their first description (Litt & Luty, 1989; Tautz 1989; Weber & May 1989) because of the high variability which makes them very powerful genetic markers. Microsatellites have proven to be an extremely valuable tool for genome mapping in many organisms (Schuler *et al.*, 1996; Knapik *et al.*, 1998), but their applications span over different areas ranging from ancient and forensic DNA studies, to population genetic and conservation/management of biological resources (Jarne & Lagoda 1996). Microsatellites offer a substantial allelic diversity at a large number of loci and are widely regarded as the most useful molecular markers available for genetic typing of individuals for kinship or large-scale population studies (Bruford & Wayne, 1993; Queller *et al.*, 1993; Jarne & Lagoda, 1996).

For instance, Baliraine *et al.* (2002) screened 24 Medfly (*C. capitata*) microsetallite markers for cross species amplification in *C. rosa, C. fasciventris* and *C. cosyra*. From the sequence analysis of the study, it was indicative that most of the Medfly based microsatellite markers could be useful for population genetic studies in the various species tested an aspect which will facilitate the tracing of the geographical origin of the colonist pest populations, assessment of their invasive potential and risk (Baliraine *et al.*, 2002). Microsatellite markers have offered an analytical tool for fruit fly invasion genetics as shown for the pest model *C. capitata* (Bonnizoni *et al.*, 2004).

KENYATTA UNIVERSITY LIBRARY

A recent study by Nardi *et al.* (2005) used microsetallite markers and mitochondrial sequences to examine the population structure and colonization history of the olive fly, *B. oleae*. The study revealed that Africa, and not the Mediterranean, is the origin of flies infesting cultivated olive, which is supported by the significantly greater genetic diversity at microsatellite loci in Africa, relative to the Mediterranean area. The results also indicate that the recent invasion of olive flies in the American region most likely originated from the Mediterranean area (Nardi *et al.*, 2005).

Microsatellite markers offer an analytic tool which may, in some cases, also differentiate between species not easily separated by morphological traits (Yu *et al.*, 2000). A study by Shearman *et al.* 2006, 12 microsatellite markers were developed from *B. papayae* which, together with markers previously isolated from *B. tryoni*, were tested on *B. papayae* and other *Bactrocera* species as species markers for the *dorsalis* complex.

2.6.3 DNA Barcoding

DNA barcoding is a novel system designed to provide rapid, accurate, and automatable species identifications by using short, standardized gene regions as internal species tags. As a consequence, it will make the Linnaean taxonomic system more accessible, with benefits to ecologists, conservationists, and the diversity of agencies charged with the control of pests, invasive species, and food safety.

DNA barcoding is a molecular technique that involves retrieval of a standard region of mitochondrial gene, Cytochrome *c* oxidase 1 (COI) at its 5' end containing approx 650

KENYATTA UNIVERSITY LIBRARY

base pairs gene to act as a 'barcode' to identify and delineate all animal life. The obtained sequence is specific for a particular species and therefore used as a 'tag' or barcode in identification of organisms in the same species. The two main goals of DNA barcoding are to assign unknown specimen to species and to enhance discovery of new species and facilitate identification, particularly in cryptic and other organisms with complex morphology (Hebert *et al.,* 2003a, b). By choosing a standard DNA fragment, the efforts of multiple research groups can be coordinated, and they are able to construct a more comprehensive library of DNA sequences than would be possible if working independently (Caterino *et al.,* 2000).

Mitochondrial DNA (mtDNA) has been widely used as a molecular genetic marker in vertebrates and invertebrates due to its characteristics, such as high-mutation rates, maternal inheritance, absence of recombination, and small molecular size (Moritz and Brown 1987). The amplification of the 650 bp region utilizes universal primers that are designed to amplify the COI gene from broad taxa. Initially, the 5' end of COI was chosen as the focal region because it is flanked by two ''universal'' primers that work for a range of metazoans (Folmer *et al.,* 1994). The need to use widely applicable primers is understandable, but examination of the DNA barcoding literature reveals that the majority of projects actually rely on taxon specific primers, rather than universal primers, in order to optimize PCR performance (Hebert *et al.,* 2004; Penton *et al.,* 2004; Barrett and Hebert, 2005; Hebert *et al.,* 2005), particularly with degraded material (Lambert *et al.,* 2005). Although many aspects of DNA barcoding have been critiqued (Sperling, 2003b; Moritz and Cicero, 2004; Will and Rubinoff, 2004; Ebach and Holdrege, 2005;

Will *et al.*, 2005), there has been only limited discussion on the decision to use a specific 600 bp fragment from the 5' end of COI as the DNA barcoding region (Erpenbeck *et al.*, 2005).

To facilitate storage, analysis and access to barcode data, the Barcode for Life Data System (BOLD-http://www.barcodinglife.org) was established as an informatics workbench. The BOLD workbench also aids in publication of DNA barcode records by assembling molecular, morphological and distributional data and bridges the gap between traditional taxonomy and bioinformatics (Ratnasingham and Herbert, 2007). DNA barcoding employs sequence diversity in short standardised gene regions to add in the large assemblies of life gene. A 648 region of the cytochrome oxidase I (COI) forms the primary barcode sequence for the animal kingdom (Herbert *et al.*, 2003a, b). BOLD was initially developed as an informatics workbench for a single, high-volume DNA barcode facility (Hajibabaei *et al.* 2005) but has evolved into a resource for the DNA barcoding community, as evidenced by its adoption for the first major barcode campaigns (birds, fishes, Lepidoptera).

CHAPTER THREE

Morphometric analysis of the new invasive fruit fly species *Bactrocera invadens* (Diptera: Tephritidae) in Africa with comparison to other *Bactrocera* species

3.1 Introduction

Species of the genus *Bactrocera* in the family Tephritidae, or 'true fruit flies', are among the most important pests of fruits and vegetables (Clarke *et al.* 2005). Because of the broad larval host range and cosmopolitan distribution, pest fruit flies are highly invasive, with adults often exhibiting a strong tendency for dispersal and larvae being readily transported to new areas via fruit movement (McPheron and Steck 1996). Recently, a pest species was recorded for the first time in Kenya (Lux *et al.* 2003) and has subsequently been found in countries across tropical Africa including: Benin, Cameron, DR Congo, Ghana, Guinea, Mali, Nigeria, Senegal, Sudan, Tanzania, Togo, Uganda, equatorial Guinea and recently, the Comoros Island (Drew *et al.* 2005; French, 2005; S. Ekesi, unpublished data). The insect was described as *Bactrocera invadens*, due to its rapid invasion of the African continent (Drew *et al.* 2005) and was rated as "a devastating quarantine pest" by the Inter-African Phytosanitary Council in 2005 (French, 2005). *Bactrocera invadens* is believed to have invaded Africa from the Indian subcontinent and was discovered in Sri Lanka after it was first reported from Africa, where it has become a significant pest of quarantine and economic importance (Mwatawala *et al.* 2004).

Taxonomic description by Drew *et al.* (2005) depicted different thoracic colourations of *B. invadens* that are morphotypes of the same pest. These colour forms have been found

in all localities and countries where the insect was collected. Awareness of how much variability exist among local populations within and between geographic regions will help in the interpretation of the biology of the fly while also contributing to a better understanding of dynamics, habitat characteristics, dispersal and distribution patterns, resources utilization and other ecological process.

As an initial effort towards management of the *B. invadens*, morphometric study has been undertaken to clarify the complexity of the pest population starting with the separation of the different colour forms and determining any morphometric differences between them and how they relate to other invasive or potentially invasive *Bactrocera* species such as *B. correcta*, *B. cucurbitae*, *B. dorsalis, B. kadiensis, B. oleae* and *B. zonata*. This will help ascertain the species limit, facilitate their reliable separation and enhance the future development of targeted management options. To achieve this objective, fourteen distances between 15 selected landmarks of all the wings were measured to characterize the shape and size of the wings for differentiation between the morphotypes and comparison with the other *Bactrocera* species (Fig. 2 & 3). Tibia length was also included in the analysis.

3.2 Materials and methods

3.2.1 Fruit fly sampling sites

Bactrocera invadens insect samples were collected from representative biogeographical regions of Kenya namely: Coast (Muhaka), Rift valley (Nguruman), Eastern (Kitui-Kasunguni) and Central (Muranga-Kaharati). Material from other countries was obtained through the African Fruit Fly Initiative (AFFI) collaborating partners. This included samples from: Ghana, Tanzania, Uganda and Zanzibar and Sri Lanka. Fruit flies were collected with methyl eugenol traps for morphometric studies and preserved in 70% alcohol; same samples were used for molecular studies. Other *Bactrocera* species included in the study are: *B. correcta*, *B. cucurbitae*, *B. kandiensis*, *B. oleae* and *B. zonata*. Table 1 shows sample data for morphometric analysis.

3.3 Morphometric analysis

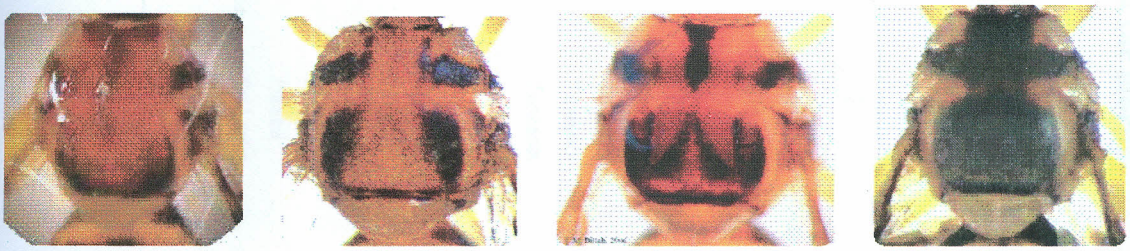
3.3.1 Source of materials

Fruit flies collected as mentioned above were separated into the four "main" morpho-type groups using the classification criteria as shown in Fig. 1:

3.3.2 Slide preparation

Specimens of the morpho-types of the *B. invadens* from the different localities preserved in 70% alcohol were used for slide preparations. The procedure followed the general processes of slide preparation with modifications according to the needs or state of the specimen. The specimens were dissected using a Leica Wild M3Z Microscope, and as follows:

Fig. 1: Morphotypes considered in the study



Type I Type II Type III Type IV

(Source: M. Billah)

Type I. Those with the pale brown thorax,

Type II. The pale forms with appearance of dark markings on the lateral margins of the scutum,

Type III. Those with increased dark markings on both the sides and middle of the thorax,

Type IV. Those with complete dark markings on the lower scutum and partial to full dark markings in the upper scutum.

Table 1: Samples for morphometric analysis

SAMPLE	COUNTRY	LOCALITY	COORDINATES	HOST PLANT	METHOD OF COLLECTION
Bactrocera invadens					
	Ghana	UBG	05° 39' 50 N, 000°11′ 15 W	Mango	ME
	Kenya	Muhaka	04° 18' 27S, 039° 29' 98E	Mango	ME
	Kenya	Nguruman	01 48' 39"S, 036 03' 28" E	Mango	ME
	Kenya	Kasunguni	01°24′ 28 S, 038°02′ 35 E	Mango	ME
	Kenya	Kaharati	00°49′ 48 S, 037°08′ 20 E	Mango	ME
	Sri Lanka	Naula	02°41′ 508 N, 80°38′ 976 E	Tropical almond	Reared
	Tanzania	Kitwilu	07° 48.043 S, 035° 40.134 E	Guava	ME
	Tanzania	Mikese	06° 44.023 S, 037° 54.330 E	Mango/Citrus	ME
	Uganda 1	Machali	01^{0}02.091N, 033^{0}96.498E	Mango	ME
	Uganda 2		00^{0} 42.587N, 032^{0}21.996E	Mango	ME
	Zanzibar	Bungi	06°14′ 850 S, 039°19′ 953 E	Mango & Citrus	ME
	Zanzíbar	Vitongoji	05°13′ 300 S, 039°49′ 402 E	Mango	ME
Bactrocera correcta	India	Bangalore	012° 57' N, 077° 37' E	-	ME
Bactrocera cucurbitae	Kenya	Nairobi	01° 13' 952 S, 036° 51' 314 E	-	LT & Cu Lure
Bactrocera dorsalis	Hawaii	Honolulu	-	Laboratory reared	-
Bactrocera kandiensis	Sri Lanka	Kandy	07°16′ 753 N, 80°35′ 731 E	Mixed fuits	ME
Bactocera oleae	Kenya	Burguret forest	00°06′ 720 S, 37° 02′ 342 E	Ex fruits	-
Bactrocera zonata	Mauritius	Reduit	-	Laboratory reared	-

- Wing and hind leg on one side (right) were carefully removed and placed in absolute ethanol for at least 15 minutes to dehydrate and transferred to clove oil for 15 min.
- The wing and leg were placed on glass slides, properly aligned, and mounted in Canada balsam thinned out in xylene. Each slide had a double mount i.e. the right wing and the right hind leg.
- The slides were then oven dried continuously for 3 – 5 weeks at 35 °C.

3.3.3 Image capture and measurement

Imaging of slide mounted specimens was done using video microscopy – Leica MZ 125 Microscope, fitted with Toshiba 3CCD camera. Images were captured using the software Auto Montage (Syncroscopy, Synoptics group, 2004) at magnification X25 for total length and width of the wing, X50 for the veins and X63 for the tibia. Measurements were taken using the program Image-Pro® Plus version 4.1 for Windows ™ (Media Cybernetics, 1999). Fourteen distances between 15 selected landmarks of all the wings and tibia lengths were measured as illustrated in fig.2 and 3.

3.4 Data analysis

Morphometric analyses were performed using the Statistical Analysis software version 8.2 (SAS Institute Inc., 2001). Principal component analysis (PCA, PROC PRINCOMP – SAS Institute, 2001) was performed on the variance-covariance matrix of the 14 wing and the tibia length variables ($\log_{10}$ transformed) to determine the effects of size and shape on the distribution of scores along the first two principal component axes (Sokal & Rohlf, 1995) and observe their distribution without constraints of prior assignment to

KENYATTA UNIVERSITY LIBRARY

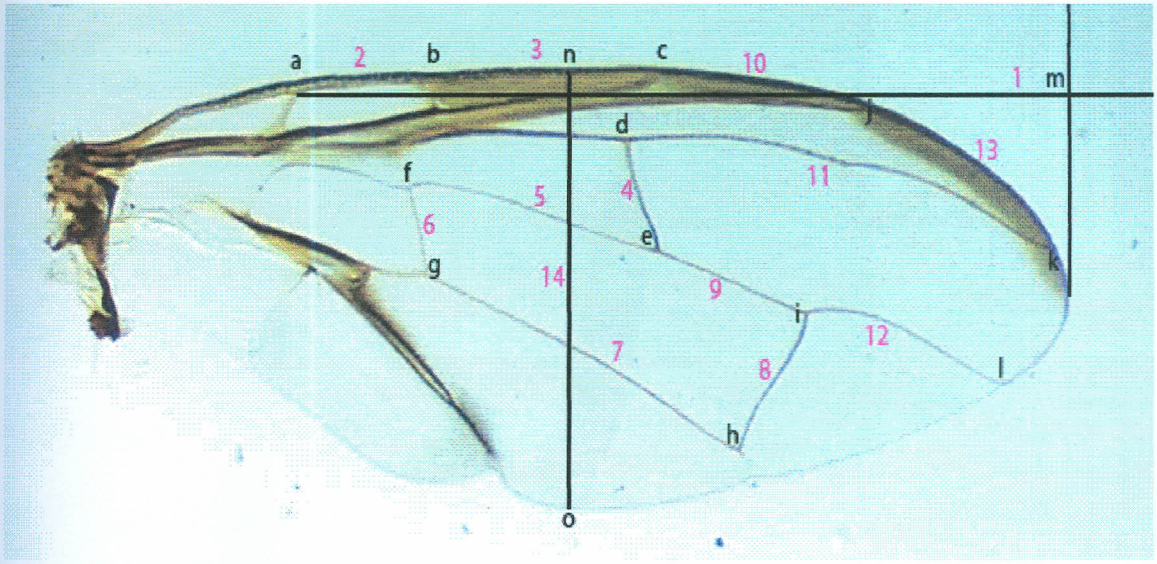


Fig. 2: Sample wing showing points of reading taken for morphometric analysis a key for identification of each wing vein

Key:

	Representation	Description
vein 1	a_m	Wing length
vein 2	a_b	Humeral break – Subcostal break
vein 3	b_c	Subcostal break – vein R1
vein 4	d_e	r – m
vein 5	e_f	Upper length of dm-cell
vein 6	f_g	Basal height of dm-cell
vein 7	g_h	Lower length of dm-cell
vein 8	h_i	Apical height of dm-cell
Vein 5_9	f_i	Upper length of dm-cell
vein 9	i_e	Upper length of dm-cell
vein 10	c_j	Vein R1 – Vein R2+3
vein 11	d_k	R_{4+5}
vein 12	i_l	M
vein 13	J_k	C
vein 14	n_o	Wing width

KENYATTA UNIVERSITY LIBRARY

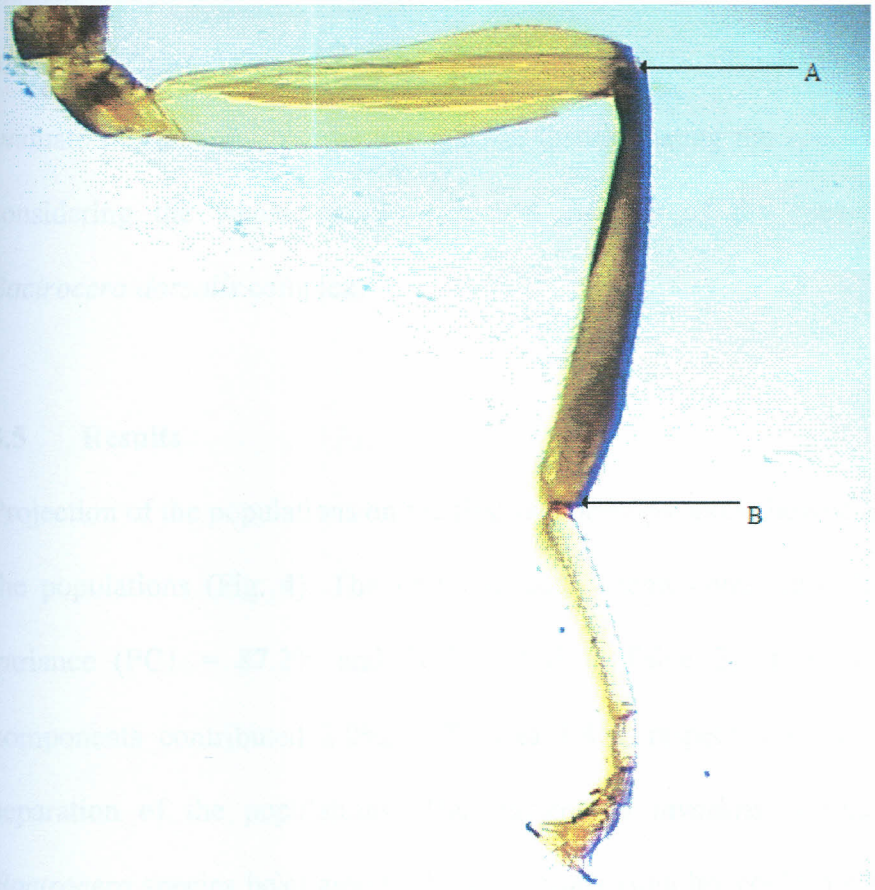


Fig. 3: Tibia points of measurement (measurement taken from point A to point B)

particular populations. The data matrix was also subjected to canonical variate analysis (CVA, PROC CANDISC – SAS Institute, 2001) to visualize shape differences and evaluate the influence of the variables in discriminating the species. This was done by considering the whole sample size and analysis of the species belonging to the *Bactrocera dorsalis* complex.

3.5 Results

Projection of the populations on the first two principal axes showed a partial separation of the populations (Fig. 4). The first two components contributed to 91.6% of the total variance (PC1 = 87.2% and PC2 = 4.4%) (Table 2). The third, fourth and fifth components contributed 2.0%, 1.7% and 1.4%, respectively, which did not improve separation of the populations. The *Bactrocera invadens* populations and the other *Bactrocera* species belonging to the *B. dorsalis* complex could not be separated by PCA (Fig. 4). However, the first two principal components separated the *B. correcta*, *B. cucurbitae*, *B. oleae* and *B. zonata* into distinct groups (Fig. 4). Likewise, the projection of the data on the first two canonical variate axes showed a similar pattern of separation (Fig. 5). The *B. invadens* populations and the other *B. dorsalis* species complex clustered together while *B. correcta*, *B. cucurbitae*, *B. oleae* and *B. zonata* distinctly separated. The first two canonical variates contributed a total of 66.2% (CV1 = 37.8% and CV2 = 28.4%) of the total variance, with the third, fourth and fifth contributing 11.3%, 10.3% and 5.9%, respectively (Table 3). The raw, standardized and total canonical structure coefficients for the canonical variate analysis of the populations are shown in Table 3. The product of the pooled within-class standard deviation and the canonical vector

coefficient for each variable is indicated as the standardized canonical coefficient (Heraty & Woolley, 1993) (Table 3) and represents the amount of change in the canonical variate source for every change in the original variable by one standard deviation (Neff & Marcus, 1980). Total sample correlations between the original variables and the canonical structure scores are represented by the total canonical structure values (Umphrey, 1996).

The largest Mahalanobis squared distance ($D^2 = 159$) is between *B. oleae* and *B. Zonata*, followed by *B. cucurbitae* and *B. zonata* (135.9), *B. kandiensis* and *B. oleae* (115.5), *B. oleae* and Sri Lanka *B. invadens* population (97.2) and *B. cucurbiate* and *B. oleae* (93.1) (Table 4). The smallest distance was between the Tanzanian and Ugandan (2.2) and Tanzania and Zanzibar (2.8) populations of *B. invadens*, an indication that these populations are closely related. The *B. zonata* sample has larger values against all the other populations analysed. The Sri Lankan *B. invadens* sample and *B. kandiensis* are very close with a distance of 6.9. When analysis was done considering only the Sri Lankan *B. invadens*, *B. dorsalis* and *B. kandiensis*, the Sri Lankan *B. invadens* was closer to *B. kandiensis* (with $D^2 = 7.2$) than *B. dorsalis* (38.5). The Sri Lankan *Bactrocera invadens* and *B. kandiensis* could not be separated by either the canonical variate or the principal component analyses (Fig. 6 & 7). The first and second principal component contributed to 83.6% and 5.5% of the separation while the canonical variate 1 and 2 contributed to 87.4% and 12.6% of the total separation of these populations respectively.

Table 2: Eigen values and weights of first two principal components, computed from log-transformed wing data of the *Bactrocera invadens* with comparison with the other *Bactrocera* species

	Weight	
Variable	Prin 1	Prin 2
Proportion of variance (%)	**87.2**	**4.4**
Eigenvalues	**0.041**	**0.002**
v1	0.23	0.08
v2	0.21	0.34
v3	0.27	-0.03
v4	0.33	-0.74
v5	0.29	-0.22
v6	0.31	-0.07
v7	0.24	0.02
v8	0.29	0.13
v9	0.26	0.34
v5_9	0.28	-0.02
v10	0.29	0.24
v11	0.22	0.12
v12	0.14	0.26
v13	0.22	0.01
v14	0.25	-0.01

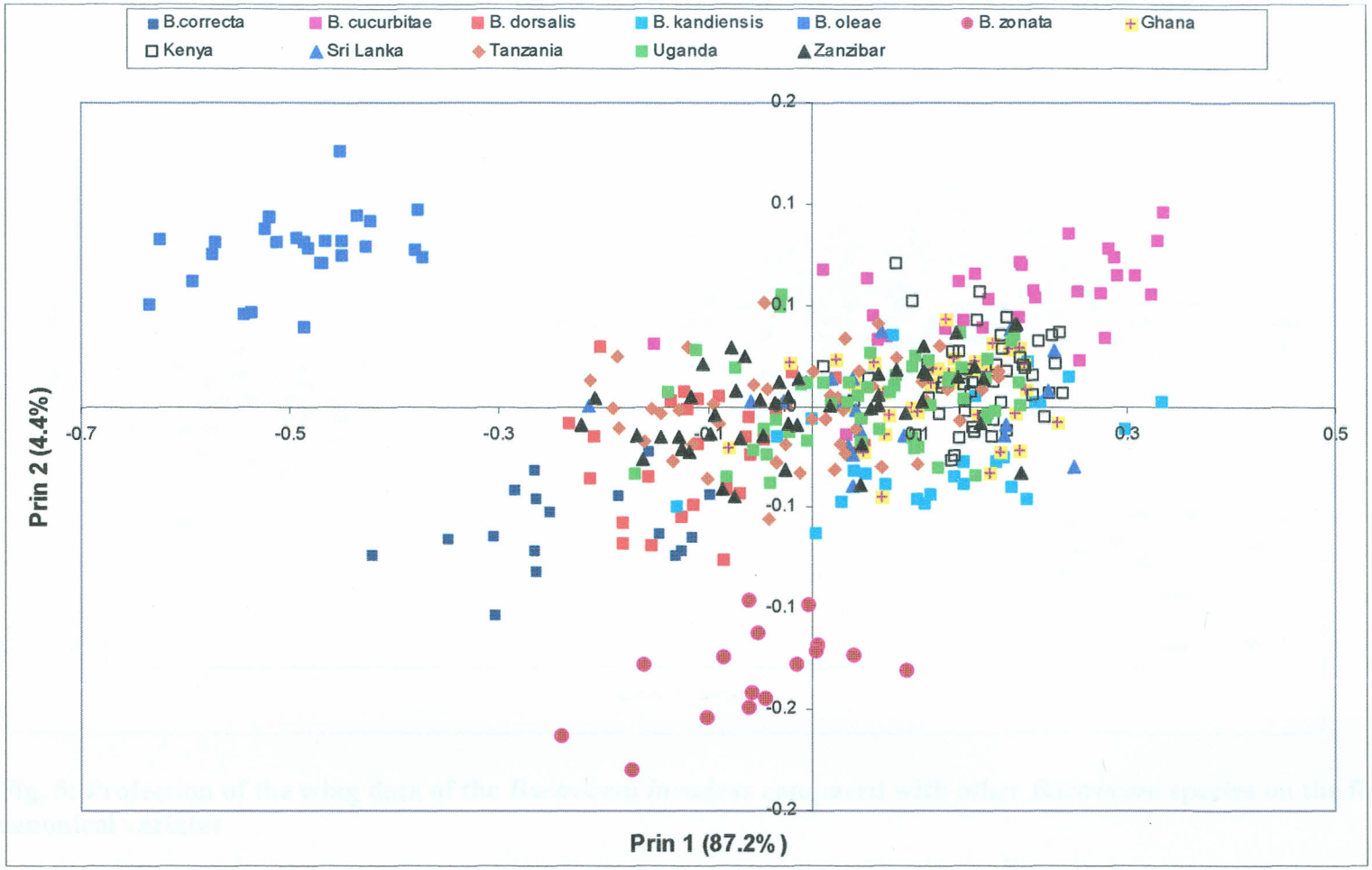


Fig. 4: Projection of the wing data of the *Bactrocera invadens* compared with the other *Bactrocera* species on the first two principal components

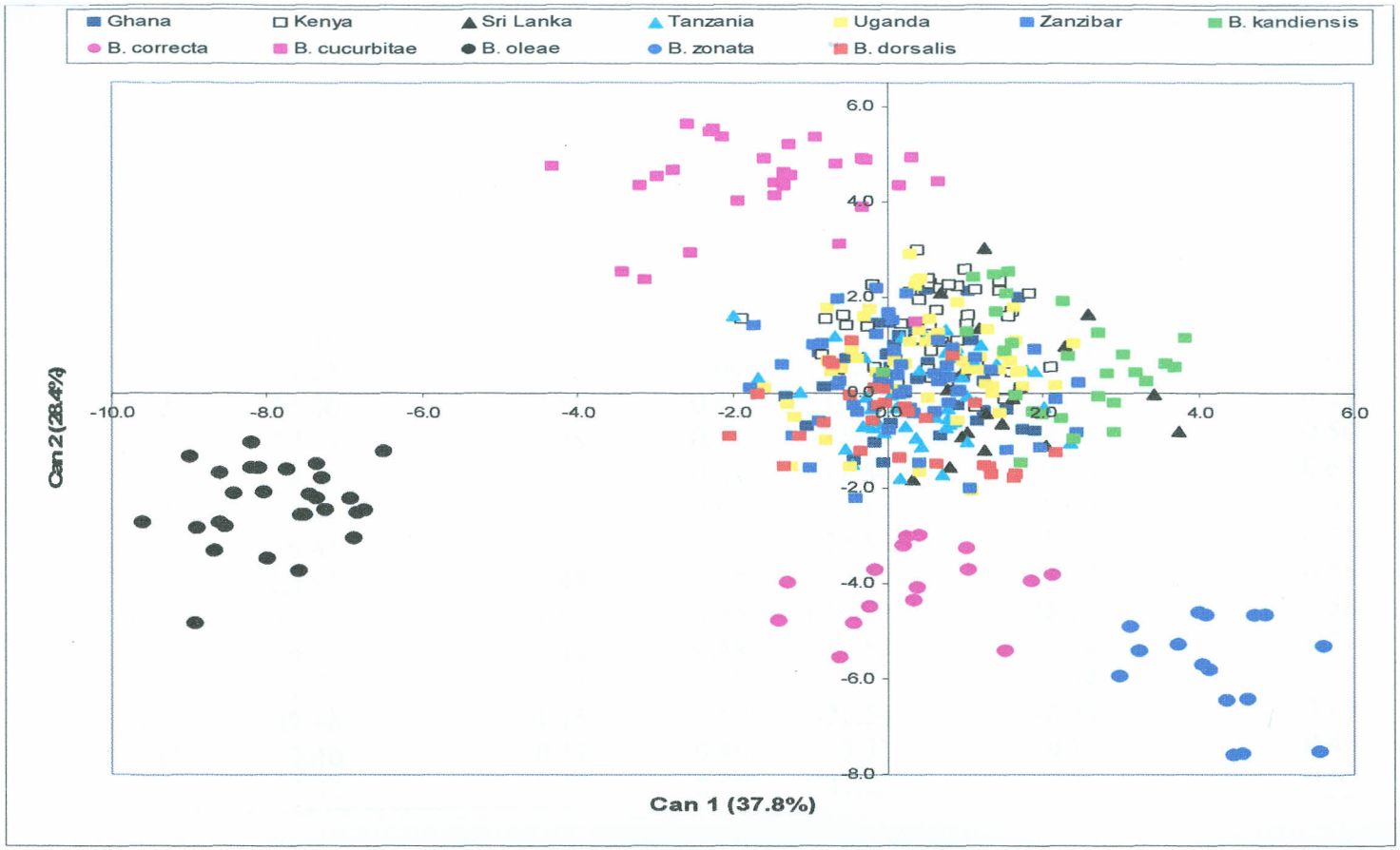


Fig. 5: Projection of the wing data of the *Bactrocera invadens* compared with other *Bactrocera* species on the first two canonical variates

KENYATTA UNIVERSITY LIBRARY

Table 3: Raw, standardised and total canonical structure coefficients for canonical variates analysis on Log-transformed wing data for the *Bactrocera invadens* and other *Bactrocera* species

	CV1 Coefficients			CV2 Coefficients		
Variable	Raw	Standardized	Total CS	Raw	Standardized	Total CS
v1	9.57	0.46	0.56	48.02	2.32	0.66
v2	-21.48	-1.05	0.25	21.11	1.03	0.78
v3	19.23	1.12	0.64	27.52	1.61	0.66
v4	35.06	2.67	0.87	-16.57	-1.26	0.28
v5	37.73	2.28	0.71	385.99	23.31	0.50
v6	9.81	0.64	0.64	12.64	0.82	0.63
v7	23.40	1.15	0.61	1.13	0.06	0.57
v8	-5.41	-0.33	0.51	25.53	1.55	0.71
v9	25.82	1.49	0.44	230.30	13.33	0.65
v5_9	-82.13	-4.65	0.65	-673.90	-38.19	0.58
v10	6.37	0.41	0.48	17.82	1.13	0.71
v11	-11.70	-0.55	0.52	7.19	0.34	0.65
v12	-32.88	-1.15	0.17	-22.54	-0.79	0.67
v13	3.40	0.17	0.46	3.31	0.17	0.49
v14	-14.80	-0.78	0.61	-41.03	-2.17	0.60

Table 4: Mahalanobis Squared Distances (D^2) between clusters representing the species/populations of the *Bactrocera invadens* and other *Bactrocera* species

Species	1	2	3	4	5	6	7	8	9	10	11	12
1. Bco	-											
2. Bcu	80.4	-										
3. Bdo	26.4	40.5	-									
4. Bka	32.8	35.2	23.9	-								
5. Bol	78.0	93.1	81.5	115.5	-							
6. Bzo	42.9	135.9	61.0	60.1	159.0	-						
7. BiGha	27.8	35.0	20.7	11.4	80.7	70.3	-					
8. BiKen	47.2	28.8	19.0	20.8	89.9	70.9	14.3	-				
9. BiSLa	32.5	40.5	29.9	6.9	97.2	61.5	10.1	26.4	-			
10. BiTzn	25.4	41.6	20.1	16.5	77.2	64.3	7.2	10.5	18.1	-		
11. BiUga	34.5	34.4	19.9	13.7	82.3	66.9	7.2	5.4	18.1	2.2	-	
12. BiZba	22.1	32.4	13.8	10.0	74.8	65.9	3.4	11.9	11.6	2.8	3.9	-

Bco-*Bactrocera correcta*, Bcu- *B. cucurbitae,* Bdo- *B. dorsalis*, Bka- *B. kandiensis*, Bol- *B.oleae*, Bzo- *B.zonata*, BiGha- *B. invadens* Ghana, BiKen- *B. invadens* Kenya, BiSLa- *B. invadens* Sri Lanka, BiTzn- *B. invadens* Tanzania, BiUga- *B.invadens* Uganda and BiZba- *B. invadens* Zanzibar.

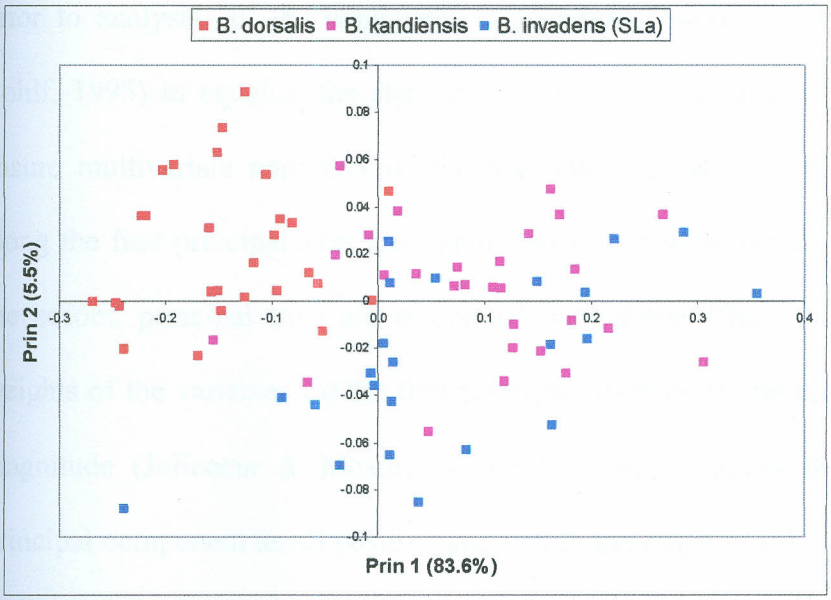


Fig. 6: Principal component analysis plot considering the Sri Lankan *Bactrocera invadens*, *B. dorsalis* and *B. kandiensis*

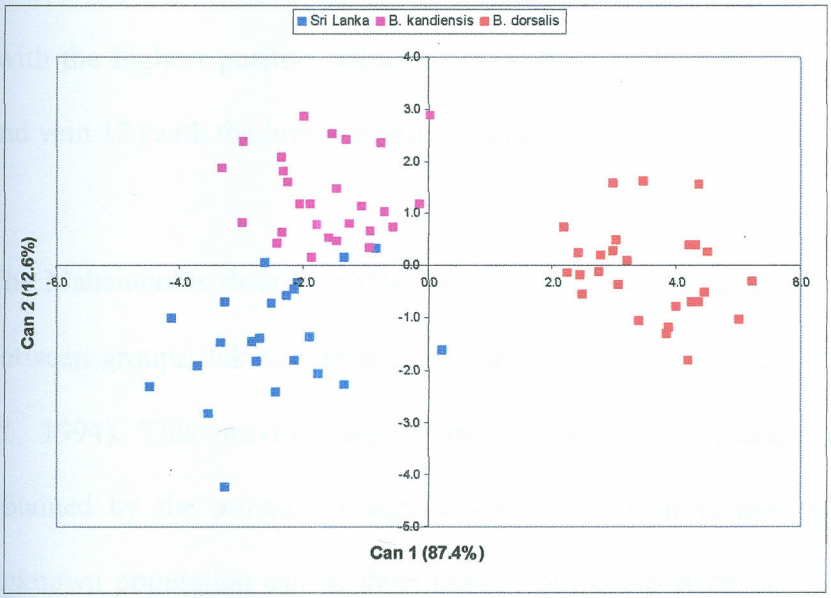


Fig. 7: Canonical variate analysis plot considering the Sri Lankan *Bactrocera invadens*, *B. dorsalis* and *B. kandiensis*

3.6 Discussion

Prior to analysis, all the wing vein measurements were log10 transformed (Sokal & Rohlf, 1995) to equalize the standard deviations across differently-sized variables and ensure multivariate normality of the data (Woolley & Browning, 1987). Separations along the first principal axes are usually associated with overall size, while those along the second principal axes are associated with shape. This is especially so when the weights of the variables for the first principal components are all positive and similar in magnitude (Jolicoeur & Mosimann, 1960). Table 2 shows the weights of the first principal component as all positive and within the range of 0.14 – 0.33, an indication of the role played by overall size in separation of the populations. Mixed magnitudes and large range values are situations where separations along the first and second principal axes are more than the simplistic association with overall size and shape (Rohlf & Bookstein, 1990). The major contributing variables were veins 4, 6, 5, 8, 10 and 5_9 (with the highest positive values of 0.33, 0.31, 0.29, 0.29, 0.29 and 0.28, respectively) and vein 12 (with the lowest positive value of 0.14).

The Mahalanobis distance values provide multivariate measures of the relative distance between groups, taking into account the variation within each one of them (Woolley *et al.*, 1994). This provides supplementary supporting evidence to the defined clusters obtained by the principal component and canonical variate plots. The identity of an unknown population can be determined from comparison of the Mahalanobis distances between the centroids of the clusters and assigning it to the population with the smallest value (Marcus, 1990) (Table 4). The D^2 showed that the *B. invadens* populations are very

KENYATTA UNIVERSITY LIBRARY

close in terms of morphometry and there is no discrimination in morphometric analysis based on the colour differences. It also indicated that the *B. kandiensis* is closer to the *B. invadens* Sri lankan sample than the *B. dorsalis ss.*

CHAPTER FOUR

Isolation and characterization of microsatellite markers in the newly discovered invasive fruit fly pest in Africa, *Bactrocera invadens* (Diptera: Tephritidae)

4.1 Introduction

Asian fruit fly pests from the genus *Bactrocera* are regarded as some of the most destructive insects of fruits and vegetables worldwide (White & Elson-Harris 1992). In March 2003, three fruit fly specimens of the genus *Bactrocera* were detected in Kenya (Lux *et al.*, 2003) and the species was described as *Bactrocera invadens* (Drew *et al.* 2005) in recognition of its rapid invasion of the African continent. It is feared that the geographical expansion of this pest and infestation could be more severe if no control programmes and efficient quarantine systems are implemented (Ekesi *et al.* 2006).

Microsatellite markers have been successfully applied to different invasive fruit fly species to infer evolutionary aspects underlying their invasive processes and to identify the routes of their colonization (Bonizzoni *et al.* 2004; Akaterawong *et al.* 2007). The isolation and characterization of 11 polymorphic microsatellite loci from the recently discovered fruit fly pest, *Bactrocera invadens* is described and provide evidence on their usefulness for the study of genetic structure and dynamics of this species.

4.2 Materials and methods

4.2.1 Samples used for the microsatellite library construction

The genomic DNA for the library construction was obtained from *Bactrocera invadens* flies from an *icipe* mass rearing strain established in 2003 from wild flies collected from mangoes in Nairobi (Kenya). Before testing for polymorphism, the markers were tested on 27 males and 18 females from the icipe mass reared strain to clear any doubt on sex linkage of these markers. The polymorphism of these loci was then tested in individual flies from two natural populations from Sri Lanka (Ranbukpitiya, 07°02'54N, 80°30'52E) and Democratic Republic of Congo (DRC, Kisantu Botanical Garden, 05°7'S, 15°5'E). Because only males are captured with methyl eugenol attractant, all the analyses were performed on male individuals, 27 from each population.

4.2.2 DNA extraction

Total genomic DNA was extracted by phenol chloroform following a protocol modified from Baruffi *et al.* (1995). Two microlitres of 20 mg/ml Proteinase K was put in eppendorf tubes containing individual flies and 200 µl of extraction buffer. The fly was crushed with a plastic pestle and the pestle rinsed with another 200 µl of extraction buffer. The resulting 400 µl of homogenate was incubated at 65 °C for 30 minutes. Fifty six microlitres of 8M Potassium acetate was added to each of the tubes mixed by inversion and incubated on ice for 30 minutes. The tubes were then centrifuged at 14000 RPM for 15 minutes at room temperature and supernatant transferred into fresh tubes. The tubes were centrifuged again for 10 minutes and supernatants transferred to fresh tubes. Two microlitres of RNase A (10 mg/ml) was added into each tube and incubated at

37 °C for 1 hour. This was followed by phenol chloroform extraction and addition of 200 µl of Ammonium acetate (7.5 M) and 800 µl of cold absolute ethanol for DNA precipitation. The tubes were then incubated at room temperature for 10 minutes and centrifuged for 15 minutes. The DNA pellet was washed with 500 µl 70 % ethanol and centrifuged for 15 minutes at 14000 RPM. The alcohol was aspirated and the pellet air dried at room temperature for 1 hour. Finally the pellet was dissolved in 50 µl of TE 1%, vortexed and stored at -20 °C. The quantity of the DNA was checked by Nano drop® ND- 1000 spectrophotometer (*Celbio*, Italy) using the program ND-1000 (Version 3.2.1). The quality of DNA was checked by 2 % Agarose gel (Appendix I).

4.2.3 Microsatellite enrichment protocol

The protocol used for the enrichment is one which was modified from a number of protocols and rewritten at the University of Pavia, Department of Animal Biology, Insect Molecular Evolutionary Biology Laboratory by Sujinda *et al.* 2006.

4.2.3.1 DNA precipitation

Six micrograms of genomic DNA was precipitated by 0.3 M Sodium acetate and the volume topped by ddH_2O to 200 µl. Cold absolute ethanol twice the reaction volume was added and the tubes centrifuged briefly, followed by freezing at -20 °C for 30 minutes. The tubes were then centrifuged at 14000 RPM for 15 minutes and supernatant discarded. The pellet was then washed with 500 µl of 70% ethanol and centrifuged again at 14000 RPM for 15 minutes. The supernatant was aspirated and discarded and the pellet air dried for 30 minutes.

4.2.3.2 Genomic DNA digestion and ligation of adaptors

Pre-mercaptoethanol KGB buffer (Appendix I) 45 µl was mixed with 5 µl of 10 mM 2-mercaptoethanol to get a 2X KGB buffer. 12.5 µl of the above 2X KGB buffer was added to 10.5 µl of ddH2O to get a total volume of 23 µl. The precipitated DNA was dissolved in the 23 µl buffer. Restriction enzyme (Appendix 1) 2µl was added to the 23 µl of the dissolved DNA for digestion. The 25 µl of restriction digest mixture was incubated at 37°C for 2 hours. To the digested DNA Oligo A (25 ng/µl) 6 µl, Oligo B (25 ng/µl) 6 µl, 10 mM ATP 6 µl and T4 DNA ligase (Fermentas, Luthiania) (3 units/µl) 2 µl were subsequently added to have a final volume of 45 µl. The mixture was incubated at 37°C for 2 hours and there after stored at -20°C overnight.

4.2.3.3 PCR amplification of the adaptor-ligated restriction digested genomic DNA

The 45 µl of the previous day mixture was thawed for 20 minutes at room temperature. PCR reaction was then set up in 0.5 µl eppendorf thin wall tubes as follow: 1/10 diluted adaptor-ligated genomic DNA (70 ng), 4.9 µl ddH_2O, 2.0 µl of 10X buffer (no premixed Mg^{2+} ions), 200 µM of 10 mM dNTP, 25 mM $MgCl_2$ and 0.5 unit Taq Polymerase (1 unit/µl). The PCR reaction was ran on an eppendorf master cycler gradient PCR machine and the conditions were as follows: 40°C for 3 minutes, 94°C for 30 seconds, 56°C for 1 minute, 72°C for 1 minute then repeating steps 2 to 4 for 26 cycles (to be altogether 27 cycles). of Oligo A primer (3.95 pmol/µl) 6.0 µl (24 pmol) was added during the first cycle of the PCR. This made a reaction volume of 20 µl. Twelve PCR reactions tubes were set up because two tubes of the PCR products had to be arrested at various PCR

cycles as following: at 20th, 23rd, 25th, 27th, 29th and 30th cycles, respectively as part of the optimization process.

The PCR product from the first tubes of each collection was visualized on a 1.5% agarose TBE system at 50 volts for 2 hours with 100 bp marker. This is to select the right cycle for proceeding with the subsequent day of the protocol. Cycle 27 and 29 were chosen to proceed to the subsequent day of the protocol after visualisation on the agarose gel. The second tubes of each collection were kept at -20°C for the enrichment step to be done on the following day of the protocol.

4.2.3.4 Enrichment of the microsatellite sequence from the RFLP product

4.2.3.4.1 Hybridisation

Before starting the protocol, the Strept-Avidin Paramagnetic Particles (SA-PMPs) were washed following the manufacturer's instructions (Promega). PCR products (20 µl) derived from the best PCR cycle (the one that yielded a DNA smear) was thawed at room temperature and 50 µl of ddH_2O and 30 µl of 20X SSC (Appendix I) was then added to the tube. The tube was then incubated in a PTC-100™ PCR machine (MJ Research) at 95°C for 10 minutes. After 10 minutes, the tube was removed from the PCR heat block at 95°C and immediately placed on ice for 2 minutes. To this denatured mix, 500 pmol of the biotinylated $(CA)_{10}$-oligonucleotide probe (125 pmol/µl) 4 µl was added and mixed gently. This hybridization reaction was then incubated at 65°C for 10 minutes. The reaction was then added to the pre-washed SA-PMPs and incubated at room temperature for 15 minutes flicking the tubes after every 2 minutes. The hybridised particles were

captured by a magnetic stand and the supernatant discarded. The particles were then washed three times with 300 µl of 0.1X SSC letting them incubate for 5 minutes at room temperature during each wash. DNA was then eluted with 25 µl ddH_2O twice to make 50 µl of eluted DNA.

4.2.3.4.2 PCR with the eluted DNA

PCR reaction was set up as follows using 0.5 ml eppendorf thin wall tubes: 4.0 µl of ddH_2O, 5.0 µl of eluted DNA, 2.0 µl of 10X buffer (no premixed Mg^{2+} ions), 400 µM of dNTP, 1.2 µl 25 mM $MgCl_2$ and 1 unit Taq Polymerase (1 unit/µl). The PCR reaction was ran on an eppendorf master cycle gradient PCR machine and the conditions were as follows: at 40°C for 3 minutes, at 94°C for 30 seconds, at 56°C for 1 minute, at 72°C for 1 minute then repeating steps 2 to 4 for 29 cycles (making 30 PCR cycles). Oligo A primer (150 ng, 3.95 pmol/µl) 6.0 µl (24 pmol) was added during the first PCR cycle. This made a reaction volume of 20 µl. Two sets of PCR reactions were set up, one set was visualized on a 1.5% agarose TBE system at 50 volts for 2 hours with 100 bp markers to select the right cycle for cloning; the other set was for cloning after selection of the right cycle. Three microlitres of this PCR product was used for cloning using TOPO TA Cloning™ Kits (Invitrogen) and as per the manufacturer's instructions. Colonies were grown in liquid media (LB media) (Appendix 1). The plasmids were extracted using 0.2 N NaOH/ 1% SDS and recombinant clones screened by digesting with EcoR1 enzyme.

4.2.4 Sequencing and primer design

A total of 200 plasmid DNA preparations derived from clones were sequenced using the BigDye™ Ready Reaction Kit (version 3.1 Applied Biosystems), with T7 forward primer (5'- TAATACGACTCACTATAGGG) and M13 reverse (5'- GTAAAACGACGGCCAGTG) on an ABI 310 genetic analyser. For every clone that contained a microsatellite locus with a suitable flanking region, a pair of primers was designed using PRIMER version 3 software (Rozen & Skaletzky, 2000). The pairs of primers which characterize each microsatellite locus were synthesized, according to the detected sequences, by Invitrogen Corporation. Preliminary PCR screening of the primers was performed on 18 females and 27 males from *icipe* strain, the same strain that was used for the construction of the microsatellite enriched library. Microsatellite primers were tested in 20µl PCR reaction containing 1x buffer, 2.5 mM $MgCl_2$, 25µM dNTP, 0.5 units Taq polymerase (Invitrogen), 10pmol of each primers and the genomic DNA. To determine the best condition of amplification, PCR reaction were carried out with 10, 20, 50 and 100ng of genomic DNA and a negative control was included for each amplification reaction. Amplifications were performed in a PTC-100™ thermocycler (MJ Research, Inc). PCR conditions were as follows: 2 min at 95°C, 39 cycles of 30 s at 94°C, 90 s at 56–61°C and 90 s at 72°C and 5 min of elongation at 72°C. The PCR products were run on a 2% agarose gel for observation of the right band size (Appendix I).

4.2.5 Genotyping

The primers obtained above and optimized were used for genotyping of the two selected populations of the *B. invadens*. Amplifications were performed in a PTC-100™

thermocycler (MJ Research, Inc) and one of the primers labeled with a florescent dye (HEX or FAM). Reactions for PCR consisted of 25 ng of genomic DNA, 1x reaction buffer, 2.5 mM $MgCl_2$, 25 µM dNTP, 0.5U *Taq polymerase* (Invitrogen) and 5 µM of each primer, one of which was labelled with a fluorescent dye at the 5' end, in a total volume of 15 µl. PCR conditions were as follows: 2 min at 95°C, 29 cycles of 30 s at 94°C, 90 s at 56–61°C and 90 s at 72°C and 5 min of elongation at 72°C. After loading on the Genetic Analyzer, PCR products were analyzed by the Genescan program (Applied Biosystem) (An example of Genescan output shown in (Appendix I). An individual allele was declared null (non amplifying allele) after at least two amplification failures.

4.2.6 Data analysis

Polymorphism of the 11 loci obtained was tested by the program Microsatellite anlayser (MSA) (Dieringer & Schlötterer 2003). The polymorphic information content (PIC) was obtained by the program Cervus version 3.0.3 (Kalinowski *et al.*, 2007). Genetic variability within populations, deviation from Hardy-Weinberg equilibrium and linkage disequilibrium was analyzed using Genepop, version 3.4 (Raymond & Rousset 1995). The program MICRO-CHECKER, version 2.2.3 (Oosterhout *et al.* 2004) was used to indicate presence of null alleles.

4.3 Results

To proceed with the microsatellite enrichment protocol, a PCR cycle with a perfect smear was obtained on the 2nd and 3rd day of the enrichment protocol. The cycle 27th and 29th

yielded smears with no bands on the 2nd day of the protocol (Fig. 8) and therefore, DNA from the two cycles was used to proceed to the 3rd day of the protocol. On this last day of the protocol, the two cycles still yielded perfect smears (Fig. 9) and cloning was done with DNA from both the cycles. Screening of the two cloned cycles was done on different plates. Prior to sequencing the positive clones (white colonies after X-Gal blue-white screening) digestion of the plasmids was done to get the size of the inserts to be sequenced (Fig. 10).

A total of 200 plasmid inserts were analyzed on an ABI PRISM 310 Genetic Analyzer (Applied Biosystems). Twenty six clones contained microsatellite sequences. Primers were designed for 11 loci (Appendix II) with suitable flanking sequence using Primer3 (Rozen & Skaletsky 2000). Preliminary PCR screening performed on 18 females and 27 males from the *icipe* strain yielded single locus amplifications and the loci were polymorphic in both sexes. This excludes any condition of sex linkage for these loci.

The 11 primers were tested to assess their polymorphism in two test populations obtained from Sri Lanka and Democratic Republic of Congo. All loci were polymorphic across the two samples (Table 5). The number of alleles per locus ranged from three to fifteen and eight loci had a polymorphic information content PIC > 0.5. Genetic variability within populations was analyzed using Genepop, version 3.4 (Raymond and Rousset 1995).

KENYATTA UNIVERSITY LIBRARY

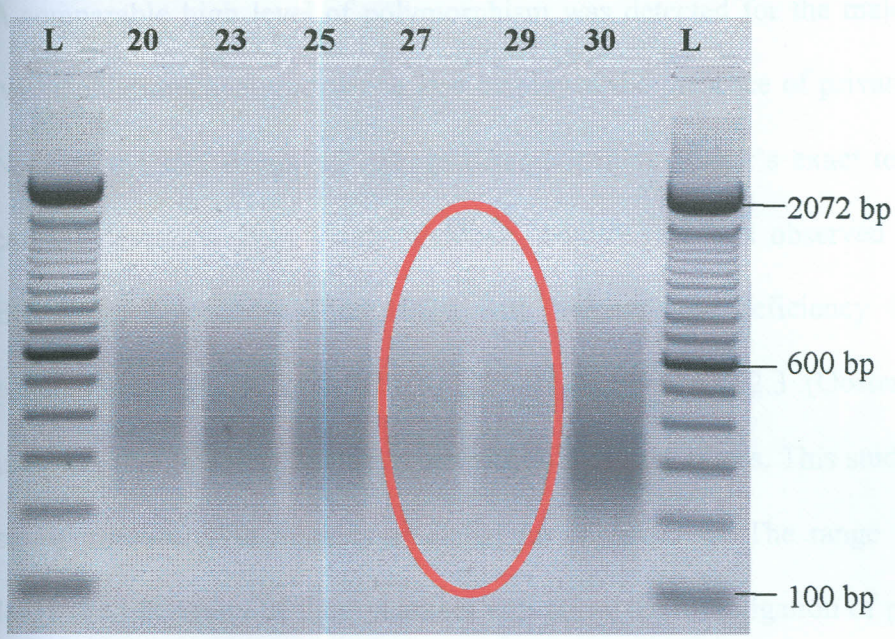


Fig. 8: Gel showing the cycles picked on the 2nd day of Microsatellite enrichment protocol

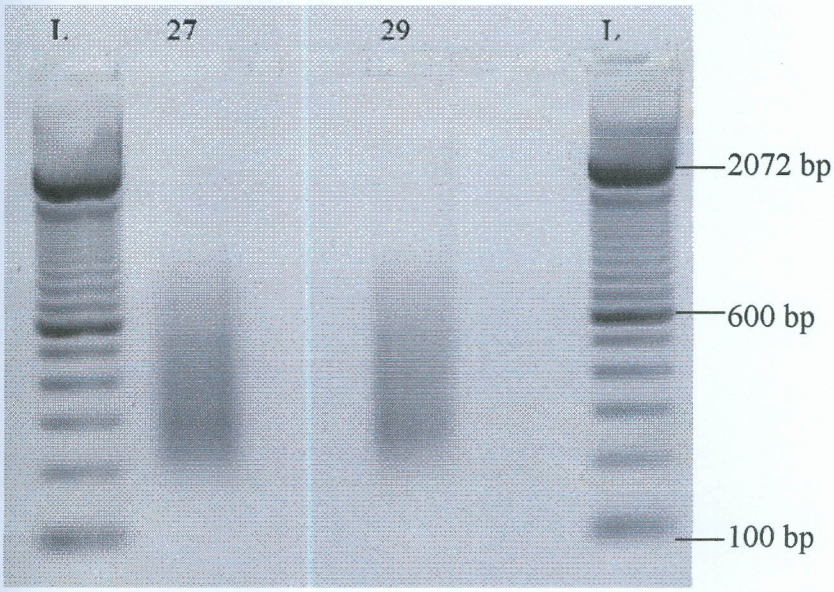


Fig. 9: Gel showing the cycles picked on the 3rd day of Microsatellite enrichment protocol

A comparable high level of polymorphism was detected for the majority of loci in the two populations analyzed which also displayed the presence of private alleles (Table 6). None of the loci displayed linkage disequilibrium (Fisher's exact test, Genepop). One case of deviation from Hardy-Weinberg equilibrium was observed at locus Bi6 after Bonferroni correction (Rice 1989) with heterozygote deficiency in the Sri Lankan population. The use of MICRO-CHECKER, version 2.2.3 (Oosterhout *et al.* 2004) indicates that this may be due to the presence of null alleles. This study provides the first set of microsatellite markers available for *B. invadens*. The range in repeat sequence length and diversity of these markers will allow the investigation of population structure and invasion history of this species.

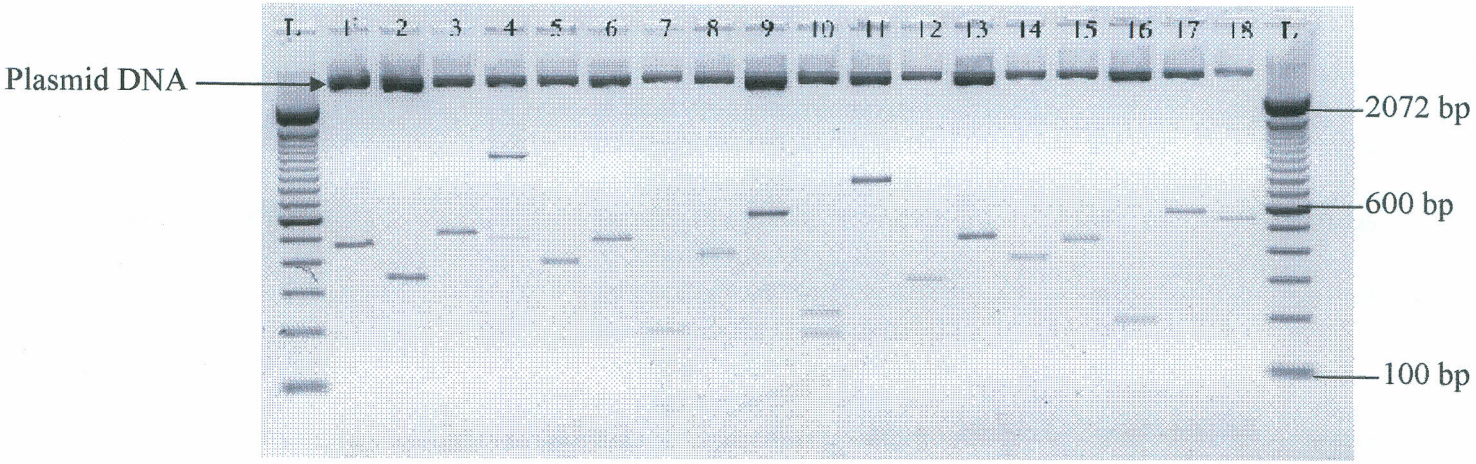


Fig. 10: A representative gel on the digestion of the plasmid DNA showing inserts of different sizes, an indication of the diversity of clones present

Table 5: Characteristics of 11 microsatellite loci from *B. invadens*. The polymorphism has been analyzed in 54 male flies, from two wild populations

Locus	Repeat motif	Primer (5'- 3')	T (°C)	N_a	Allele size range (bp)	PIC
Bi1	$(CA)_9TT(CA)_2$	HEX CTCTTGACACTGGCCTCGTT	58	12	129-159	0.833
		R: GTATGGCCGGAGACATCAGT				
Bi2	$(CA)_7$	FAM GCACTCACTCAACCTTAGACG	59	10	140-149	0.811
		R: GCACTTGAATTGTGCGAAGT				
Bi3	$(TG)_5CTG$	FAM CGCGAATTTCAAGCATTTTT	60	3	130-132	0.103
		R: GGTCTTAAGGCCAAGCACAA				
Bi4	$(CT)_6GCT$	HEX GCACTCGCATGCTTGTAGTC	59	4	123-126	0.683
		R: CCGGTTTTGTGCGAAAAG				
Bi5	$(CA)_5GA(CA)_2$	HEX GCCAGTCAGTGTCTCGTCAA	60	15	112-133	0.829
		R: AGCGAGTTTGTTTTCGGTGA				
Bi6	$(CA)_3CC(CA)_3$	FAM GCGACAAGTTCGACACAAAA	60	8	90-101	0.707
		R: TACTGATTGTGCCGTGTGCT				
Bi7	$(CA)_5A(CA)_2$	FAM CTCGCTCTTCATTCAATCCA	58	5	107-118	0.383
		R: CGACACGTTAAGTGGCAAAA				
Bi8	$(AC)_5AT(AC)_2$	FAM ACAAGTGCAGCAAAGACACG	61	7	118-138	0.569
		R: ATCACATCATGAGGCGTTCA				
Bi9	$(TG)_2TA(TG)_6$	HEX GCGCTGCTCGTAAACATCTA	61	9	92-108	0.721
		R: GGGCAAACACTTGGATTCAC				
Bi10	$(TG)_3AT(TG)_8$	FAM ATCGAGCAGATCACTGAGCA	56	11	142-155	0.757
		R: CCGCGAGTAGCAAATCTTTC				
Bi11	$(TC)_3T(TC)_3$	HEX TGGGTTCACCGTCCTTTAAT	60	3	140-147	0.208
		R: GCCCATAGACATCCAGGGTA				

HEX and FAM: Fluorescent dye labelled at the 5' end of the forward primer; T_a: annealing temperature; N_a: number of alleles per locus; PIC: polymorphic information content (Kalinowski *et al.* 2007).

Table 6: Diversity statistics for 11 microsatellites loci of *Bactrocera invadens* from two natural populations

	Natural Populations									
	Sri Lanka (n = 27)					DRC (n = 27)				
Locus	N_a	H_O	H_E	n_O	n_p	N_a	H_O	H_E	n_O	n_p
Bi1	12	0.85	0.88		6	6	0.67	0.75		0
Bi2	10	0.54	0.83		5	5	0.72	0.75		0
Bi3	3	0	0.14		1	2	0	0.07		0
Bi4	4	0.78	0.74		0	4	0.48	0.70		0
Bi5	15	0.81	0.91		11	4	0.41	0.68		0
Bi6	6	0.20	0.75*	0.34	2	6	0.58	0.67		2
Bi7	5	0.56	0.49		2	3	0.22	0.31		0
Bi8	6	0.59	0.57		3	4	0.63	0.62		1
Bi9	8	0.74	0.75		4	5	0.85	0.74		1
Bi10	7	0.63	0.83		4	7	0.63	0.71		4
Bi11	3	0.22	0.26		1	2	0.19	0.17		0
Mean	7.18	0.54	0.65			4.36	0.49	0.56		
SD	3.82	0.28	0.26			1.63	0.26	0.25		

n: number of males tested; N_a: number of alleles per locus; H_O and H_E are observed and expected heterozygosities, respectively; * deviations from HWE at $P < 0.05$ after Bonferroni correction; n_O: null alleles frequency and n_p: number of private alleles.

KENYATTA UNIVERSITY LIBRARY

4.4 Discussion

The microsatellite enrichment protocol employed in this study has been successfully applied previously on the *Bactrocera dorsalis sensu stricto*. These markers were used to infer its colonisation and invasion history in Asia and the pacific region (Aketarawong *et al.*, 2007). This study provides the first set of microsatellite markers available for *B. invadens* that could be used to study the genetic structure and the invasion history of the *B. invadens* in Africa.

In the *B. invadens*, out of 200 recombinant clones sequenced, only 26 (13%) contained microsatellite loci. This outcome might indicate that the genome of the *B. invadens* is not rich in microsatellite loci. While in other *Bactrocera* species, for example *B. dorsalis s.s* (Aketarawong *et al.*, 2006) out of 87 recombinant clones sequenced 39 contained microsatellite loci (44.8%) and in *B. cacuminata* (Song *et al.*, 2006) out of 125 positive sequenced clones, 75 had microsatellite loci (60%). The microsatellite repetition more frequently observed in *B. invadens* is the $(CA)_n$, which is also observed in the *B. cacuminata* (Song *et al.*, 2006), *B. dorsalis ss* (Aketarawong *et al.*, 2006) and *B. oleae* (Augustinos *et al.*, 2002). In other Diptera such as *Bactrocera tryoni* (Kinnear *et al.*, 1998), *B. papayae* (Shearman *et al.*, 2006), *C. capitata* (Bonizzoni *et al.*, 2000), *D. melanogaster* (Schung *et al.*, 1998a, b; Bachtrog *et al.*, 1999), *D. simulas* (Hutter *et al.*, 1998) and *Anopheles gambiae* (Zheng *et al.*, 1996) the most frequent dinucleotide repeat is the $(GT)_n$.

Although *B. invadens* was first detected in 2003 (Lux *et al.*, 2003a) and described as a separate species in 2005 (Drew *et al.*, 2005), it has already acquired a great economic significance in Africa. On this basis, the study of the invasive process of this fly is of fundamental importance to be able to interpret its history and trace its entry into Africa and map the pattern of its spread. The use of molecular tools to reconstruct the history of invasions and to identify the source of introduced populations may also facilitate the identification of effective biological control agents for introduced species (Tsutsui *et al.*, 2001).

CHAPTER FIVE

Uncovering the tracks of a recent and rapid invasion: the case of the fruit fly pest *Bactrocera invadens* in Africa

5.1 Introduction

In 2003, an unknown *Bactrocera* species, was detected in Kenya (Lux *et al.*, 2003) soon after the completion of a programme of monthly fruit collection carried out from 1999 to 2003 (Copeland *et al.*, 2004). The insect was described as *Bactrocera invadens* (Drew *et al.*, 2005) and has been rated as "a devastating quarantine pest" by the Inter-African Phytosanitary Council, in 2005 (French, 2005). Within two years of its discovery in the coastal region of Kenya, the species has been recorded in several countries on the African mainland (Mwatawala *et al.*, 2004; Drew *et al.*, 2005; Ekesi and Billah, 2007). It is now known to be present in tropical Africa from Senegal to Mozambique, as well as in the Comoro Islands in the Indian Ocean.

Bactrocera invadens is believed to belong to the *Bactrocera dorsalis* complex of tropical fruit flies (French, 2005). The *B. dorsalis* complex comprises more than 75 species largely endemic to South-East Asia (Drew & Hancock, 1994) with undescribed species remaining in collections (Lawson *et al.* 2003). Only a few species of *Bactrocera* have become established outside their native oriental range, namely *Bactrocera cucurbitae* (Coquillet) in eastern Africa and Hawaii, *Bactrocera latifrons* (Hendel) in Hawaii, *Bactrocera oleae* (Gmelin) in California and members of the *B. dorsalis* species complex in both Hawaii and Surinam (White & Elson-Harris, 1992; Rice *et al.* 2003; Clarke *et al.*

2005). The invasion of *B. invadens* in Africa adds to this list and compounds the existing fruit fly problem in Africa.

The timing and the pathway of invasion by *B. invadens* in Africa are unknown. The fact that the first historical records of this pest in Africa were from East coast, may indicate that this area was the port of entry of *B. invadens* in Africa, but this hypothesis has not been tested. Moreover the native range is not well defined: it has been suggested that it ranges from Sri Lanka to the Southern Indian sub-continent from where the species may have invaded Africa (Mwatawala *et al.,* 2004; Drew *et al.,* 2005). In recent years, the detrimental effects of this invasive species are stimulating studies to define its ecological niche and the invasion potential (Ekesi *et al.,* 2006; De Meyer *et al.,* 2009; Mwatawala *et al.,* 2006). However, due to the "novelty status" of this fruit fly as disperser/invader, no data are yet available on the genetic diversity and on the degree of co-ancestry among the African and between them and the supposed native populations from Southern India and Sri Lanka. As a consequence, no inferences, based on genetic data, were possible about the invasion route of this pest.

Using microsatellites markers (SSRs) previously characterised in the course of this study (Khamis *et al.,* 2008), the level of genetic diversity and the extent of the common ancestry among several African populations collected across the actual invaded area in tropical Africa from the East to the Western regions were evaluated. Given the short time span between the detection of the invasion and the observations of this study, the results highlight the dynamic aspects of the invasion process after the introduction. The same

data, correlated with the history of the first records of this pest in different countries, provide the key to infer a possible route of invasion in the African mainland. Moreover a sample population from Sri Lanka is tested to confirm the Asian origin of this pest.

5.2 Materials and method

5.2.1 Sample collection

Samples of *Bactrocera invadens* (males) were collected by Methyl Eugenol traps (except the Nguruman samples which were reared from mango fruit (*Mangifera indica*)) from 13 localities in four regions representing the species range, namely East Africa, Central Africa, West Africa and South East Asia (Table 7). East Africa includes samples from Kenya (three localities), Tanzania (one locality) and Zanzibar (one locality), and Uganda (two localities); Central Africa with Democratic Republic of Congo (one sample); West Africa with Nigeria (one locality), Benin (one locality) and Ghana (two localities); South East Asia with one locality in Sri Lanka. All samples were preserved in 95% ethanol prior to genetic analysis, and voucher specimens are available at both the University of Pavia and *icipe*. Details of collection are summarized in Table 7.

5.2.2 DNA extraction

In each of the above populations, 27 insects were picked from representatives of the morphotypes of the *B. invadens* and each individual fly used for DNA extraction following the method of Baruffi *et al.* (1995) as described in chapter four.

KENYATTA UNIVERSITY LIBRARY

Table 7: Collection data of *B. invadens* populations used in the population genetic studies

Region	Sample name*	First record	Sample site	Date of collection	Coordinates	Elevation (m)
Asia						
South Asia	Sri Lanka	1993*	Naula	2005	07°41'N, 80°39'E	299
Africa						
East Africa	Kenya 2	2003**	Nguruman Rift Valley	2006	01°44'S, 35°58'E	700
	Kenya 3		Kitui-Kasunguni	2005	01°21'S, 38°00'E	1010
	Kenya 4		Muranga-Kaharati	2005	00°49'S, 37°13'E	1602
	Tanzania	2003†	Morogoro SUA	2005	08°14'S, 36°57'E	507
	Zanzibar	?	Bungi	2005	06°14'S, 39°19'E	14
	Uganda 1	2004*	Kawanda	2007	00°49'N, 31°55'E	1172
	Uganda 2		Kisule	2005	00°44'N, 32°31'E	1153
	DR Congo	2005	Kisantu Botanical Garden	2006	05°10'S, 15°08'E	559
West Africa	Nigeria	2003‡?	Zaria-Kaduna	2005	11°06'N, 07°42'E	641
	Benin	2004*	Monts Kouffé	2005	08°42'N, 02°05'E	447
	Ghana L	2005*	Legon, Ghana University	2005	05°39'N, 00°11'W	89
	Ghana M		Mampong-Akwapem	2005	07°04'N, 01°25'W	412

* Drew et al. 2005; ** Lux et al. 2003; † Mwatawala et al. 2004; ‡ Umeh et al. 2008

5.2.3 Microsatellite genotyping

All samples were scored at eleven polymorphic microsatellite loci developed for *B. invadens* previously described (Khamis *et al.*, 2008). Primer sequences and the methods used for DNA amplification, electrophoresis and allele scoring using an automated ABI PRISM 310 Genetic Analyzer (Applied Biosystems) were as reported above. An individual was declared null (not amplified at a locus) only after at least two amplification failures.

5.3 Data analysis

5.3.1 Genetic diversity and differentiation

The following genetic diversity estimates for each sample were measured as averages over loci with the software Genepop, version 3.4 (Raymond and Rousset 1995): mean number of alleles (n_a), number of rare alleles (n_r), i.e. with frequency less than 0.05, number of private alleles (n_p), frequency of private alleles (A_p), observed heterozygosity (H_O) and expected heterozygosity (H_E). Gene diversity (H_S), allelic richness (R_S) and the inbreeding index (F_{IS}) estimates were assessed using FSTAT 2.9.3.2 (Goudet 2001). The program Genepop was also used to test for linkage disequilibrium between pairs of loci in each population (100 batches, 1000 iterations per batch) and for deviations from Hardy–Weinberg (HW) equilibrium at each locus/population combination using Fisher's exact tests. Bonferroni correction was used for all tests involving multiple comparisons (Rice 1989).). The frequency of null alleles was calculated using Brookfield estimation (Brookfield, 1996) in Micro-Checker 2.2.3 (van Oosterhout *et al.*, 2004, 2006). Micro-Satellite Analysis (MSA) software, version 4.05 (Dieringer & Schotterer 2003) was

applied to study the degree of genetic differentiation between pairs of populations, in term of pairwise Fst values (Weir & Cockerham 1984). The statistical significance of each value was assessed by the comparison of the observed value with the values obtained in 10,000 matrix permutations.

5.3.2 Structure analyses

The genetic structure of the thirteen *B. invadens* populations was inferred with the program STRUCTURE version 2.1 (Pritchard *et al.* 2000) using the admixture model with correlated gene frequencies (the "F model') (Falush *et al.* 2003). The admixture model with correlated frequencies allows for individuals to have mixed ancestry, with alleles frequencies in the populations likely to be similar because of migration or shared ancestry. The optimum number of clusters (K) was determined by using the admixture model, and testing for different possible numbers of populations (K between 1 and 13 populations were considered Additional parameters used were: different values of F_{ST} for different subpopulations, prior F_{ST} mean 0.01, standard deviation 0.05 and constant lambda valued at 1. Convergence of the Markov chain Monte Carlo (MCMC) was obtained for burn-in and MCMC run lengths of 50 000 and 100 000, respectively. The most likely number of clusters, given the data was chosen considering the number of clusters where the largest difference in log likelihoods was observed (ΔK, Evanno *et al.* 2005). To indentify the most likely number of sub-populations (K) among individuals we used ΔK, the maximum second order rate of change of posterior probability given the data, Pr(X/K), standardized by the standard deviation of Pr(X/K) (Evanno *et al.*, 2005).

The mean and standard deviation of Pr(X/K) and ΔK were calculated from 10 replicate analyses for each value of K.

5.3.3 Principal coordinate analysis

The principal coordinate was calculated on Nei's genetic distances between populations. Nei's genetic distance matrix was built using the program Microsatellite Analyzer (Dieringer & Schlötterer 2003), and principal component analysis was performed under GENALEX (Peakall and Smouse 2005).

5.3.4 Analysis of Molecular Variance

To complete the clustering analysis, a locus by locus analysis of molecular variance (AMOVA) was performed using Arlequin software, version 3.1 (Excoffier *et al.*, 2005). AMOVA was employed to assess the component of genetic diversity attributable to variance between and within four different population groupings. Estimates of population divergence were computed using variance in allelic frequencies (F-stat: Weir and Cockerham, 1984). Significance for AMOVA analysis was ascertained using 10,000 permutations.

5.3.5 Population Bottleneck

The program BOTTLENECK (version 1.2.02; Cornuet & Luikart 1996) was used in the attempt to recognize the effect of a recent bottleneck in the African history of *B. invadens*. One of the assumptions of the method is that allele frequency distribution result from equilibrium between mutation and genetic drift. Recent bottlenecks provoke a shift

away from an L-shaped distribution of allele frequencies, to one with fewer alleles in the low frequency categories. Under specific assumptions relative to the SMM mutation model, the methods imply the notion that shrinking populations reduce allelic diversity faster than heterozygosity or gene diversity (Nei *et al.* 1975). An excess of observed gene diversity relative to the gene diversity expected from the number of alleles observed in the sample may indicate a population size reduction, whereas a deficit of observed gene diversity may indicate that the population is growing. Two mutation models, considered appropriate for microsatellites (Piry *et al.* 1999), were applied: the strict Stepwise Mutation Model (SMM), and the Two-Phase Model (TPM). For the TPM, a model that includes both 90% single-step mutations and 10% multiple step mutations was used. Significant deviations in observed heterozygosity over all loci were tested using a nonparametric Wilcoxon test.

5.3.6 Population assignments and exclusion

The program GENECLASS 2.0 (Piry *et al.*, 2004) was used to assign or exclude reference populations as possible origins of individuals, on the basis of multilocus genotypes. The program calculates, for each individual of a population, the probability of belonging to each other reference population or to be a resident in the population where it was sampled. The standard criterion described by Rannala & Mountain (1997), which applies Bayesian statistics to compute assignment probabilities, was used. The Monte Carlo resampling method (Paetkau *et al.*, 2004) is aimed to the identification of accurate exclusion/inclusion critical values: the results are based on 10,000 simulated genotypes for each population and on a threshold probability value of 0.01.

5.3.7 Phylogenetic analysis

Phylogenetic trees of the *B. invadens* populations were constructed using the PHYLIP package (Felsenstein 2005). The genetic distance by Cavalli-Sforza (D_C) (Cavalli-Sforza & Edwards 1967), by Nei (D_A) (Nei *et al.*, 1983) and the one based on the proportion of shared alleles (D_s) (Bowcock *et al.* 1994) used for tree derivation, were calculated by the MSA software. These three distances (D_C, D_A and D_s) were used to construct neighbour-joining (NJ) trees after 1000 bootstrap resampling of the allele frequency data, and subsequently the graphics drawn and viewed by TREEVIEW 1.6.5 (Page, 1996).

5.4 Results

Genetic diversity

In a total of 351 flies coming from 13 locations genotyped at 11 microsatellite loci, we identified 105 alleles. All the loci were polymorphic in terms of allele number and allele size range. The mean observed heterozygosity (H_O) and expected heterozygosity (H_E) per population over all loci ranged from 0.39 to 0.56 and 0.49 to 0.65, respectively. After sequential Bonferroni correction (Rice 1989) for multiple comparisons, exact tests revealed that 24 of the 143 population/locus combinations showed significant departures from Hardy-Weinberg equilibrium, involving 9 of the 11 loci, all exhibiting heterozygote deficiencies. However the locus / population combination that were not in Hardy-Weinberg equilibrium were not concentrated in any population or at any locus. No significant pairwise linkages were detected between the 11 loci (Fisher's exact test, Genepop).

KENYATTA UNIVERSITY LIBRARY

For each population the estimated genetic variability across the 11 loci is summarized in Table 8. The Asian sample of Sri Lanka presents the highest values of genetic variability, with 7.18 mean number of alleles per locus, 16 private alleles at a mean frequency of 0.04, and with an high value of genetic diversity (H_S = 0.65). Compared with Sri Lanka, lower variability values are observed in the African regions, where the Western populations display a slightly higher level of genetic diversity: the high number of rare alleles which characterizes West African regions, especially Nigeria, contributes highly to the gene diversity in West Africa. An average deficiency of observed heterozygosity is observed in some of the samples; one cause could be the presence of null alleles. The null allele frequency estimates vary between 0 and 0.12. Nevertheless the deficiency of observed heterozygosity can be related to a Wahlund effect considering that higher values of F_{IS} are observed in the West Africa, where a high number of rare alleles are also observed.

Population structure

Differentiation among populations, as measured by the fixation index F_{ST} (Table 9) represents the first approach to the population structure of the *B. invadens* samples. Pairwise F_{ST} values were generally low, with F_{ST} ranging from 0.015 to 0.134. There were no out-of-range high values observed between African populations and Sri Lanka. Low values of differentiation are observed among eastern African populations: here four of the eight values were found not to be significantly different from 0. In West Africa, Nigeria has the highest values of F_{ST} with all eastern African populations, but this population shows a certain degree of isolation also among the western populations.

Within the western African regions the only not significant F_{ST} value (0.015) has been detected between Benin and Ghana 1, since the other three non significant values regard the two Ghana samples with eastern populations.

In a principal component analysis, the first three principal coordinates accounted for 27.63%, 23.55% and 20.15% of the variance, respectively (Fig. 11). The first and second principal coordinates highlights the differentiation of the Sri Lanka from the African populations that clustered into two. Likewise, the two axes separated the Nigerian population from the two African clusters. The first cluster consists of Benin, clustering together with Ghana 1, Kenya 2, Tanzania 1, Uganda 1 and Uganda 2. Second cluster consists of Kenya 3 sharing the cluster with DRC 1, Ghana 2, Kenya 4 and Zanzibar 1. The lack of strong regional structure among our samples was also evidenced by AMOVA (Table 10). Up to (not significant) 5% of genetic variance resulted from partitioning among groups in the four tests performed, where the bulk of genetic variation was within groups and among individuals within populations. When Sri Lanka is not considered, and only African samples are compared, the eastern and western African samples show no proper molecular variance.

Results of structure analysis showed a $\Delta K = 4$, an indication that the 13 populations can be described within 4 hypothetical genetic subdivisions defined as clusters (K). The average value of ancestry probabilities (Q) of each population in the clusters is illustrated in Table 11. While Sri Lanka and Nigeria have the highest coancestry in two separate clusters, i.e. the fourth (Q=0.941) and the third (Q=0.837) cluster respectively, it appears

that the ancestry of African flies is distributed in cluster 1 and 2, without a clear cut separation of eastern from western flies.

In cluster 1, the highest coancestry values are dominated by Kenya-3 (Q=0.916) and Kenya-4 (Q=0.847).The same cluster shares coancestry with DRC (Q=0.791), Ghana 2 (Q=0.711), Zanzibar (Q=0.700) and with Tanzania and Uganda with the lowest probabilities. In cluster 2, the Ghana-1 sample has the highest coancestry value (Q=0.930), and we have the clustering of Kenya-2 (Q=0.880), Uganda samples (Q=0.86-0.700), Tanzania (Q=0.637) and Benin (Q=0/618). Eastern flies are therefore more or equally represented, with respect to the western ones, in these two African clusters. Admixed lineages are present in Eastern and Western populations and, among the others; Benin and Zanzibar are the samples which had far greater number of admixed individuals, sharing some degree of ancestry also with Nigeria. The distributions of all countries in their respective clusters are presented on pie charts in a map (Fig. 12).

Assignment rates

The assignment test of GeneClass2 is a direct method to ascertain population membership of individuals. The diagonal values of the assignment matrix in Table 12 indicate the average probability with which individuals are assigned to the corresponding reference population: all self-assignment probability values range from 0.396 (Tanzania) to 0.507 (Kenya-K). As the program computes the probability that the multilocus genotype of each individual can be encountered in a given population, the very low probability values

Table 8: Genetic variability estimates in field –collected samples of *B. invadens* from different geographical regions

	Population	n_a	n_r	n_p	A_p	$R_S \pm SD$	$H_S \pm SD$	H_O	H_E	F_{IS}	A_n
South Asia	Sri Lanka	7.18	35	16	0.04	7.07±3.72	0.65±0.25	0.54	0.65	0.17*	0.07
East Africa	Kenya 2	3.73	7	0	0	3.70±1.58	0.49±0.27	0.39	0.49	0.20*	0.07
	Kenya 3	4.00	8	0	0	3.96±1.59	0.56±0.24	0.56	0.56	0.001	0.00
	Kenya 4	3.91	5	1	0.04	3.89±1.64	0.51±0.31	0.42	0.51	0.17*	0.06
	Tanzania	4.18	9	0	0	4.13±2.17	0.52±0.32	0.45	0.52	0.15*	0.05
	Zanzibar	5.00	14	2	0.02	4.96±2.25	0.61±0.23	0.53	0.61	0.12	0.05
	Uganda 1	4.55	15	3	0.02	4.47±2.46	0.53±0.30	0.43	0.53	0.19*	0.07
	Uganda 2	4.27	7	0	0	4.26±2.05	0.53±0.28	0.47	0.53	0.10	0.04
	DR Congo	4.36	12	0	0	4.31±1.60	0.56±0.25	0.49	0.56	0.13	0.04
West Africa	Nigeria	5.45	21	2	0.03	5.35±2.08	0.64±0.15	0.46	0.63	0.28*	0.10
	Benin	4.82	10	1	0.02	4.78±2.31	0.60±0.24	0.41	0.60	0.32*	0.12
	Ghana 1	4.27	10	1	0.04	4.23±1.44	0.55±0.23	0.43	0.55	0.21*	0.08
	Ghana 2	4.45	8	0	0	4.41±1.69	0.58±0.27	0.45	0.58	0.22*	0.08

n_a mean number of alleles; n_r number of rare alleles (frequency<0.05); n_p number of private alleles; A_p mean frequency of private alleles; R_S allelic richness and SD is the standard deviation; H_S gene diversity; H_O mean observed heterozygosity; H_E mean expected heterozygosity; F_{IS} the inbreeding index, the asterix indicates the statistical significance after Bonferroni correction at $P<0.05$; A_n mean frequency of null alleles $[(H_E - H_O)/(H_E + 1)]$ (Brookfield 1996)

Table 9: Pairwise F_{ST} divergence between 13 different geographical populations of *B. invadens* as derived from Microsatellite Analyser (Dieringer and Schlotterer 2003). Some of the values are not significantly different from 0 (n.s)

Population	1	2	3	4	5	6	7	8	9	10	11	12	13
01 Sri Lanka		0.134	0.122	0.099	0.076	0.091	0.107	0.097	0.078	0.125	0.083	0.117	0.068
02 Kenya 2			0.098	0.065	0.029 n.s	0.107	0.035	0.067	0.085	0.125	0.037	0.018 n.s	0.073
03 Kenya 3				0.056	0.085	0.072	0.106	0.125	0.057	0.129	0.094	0.107	0.080
04 Kenya 4					0.029 n.s	0.056	0.050	0.063	0.037	0.110	0.062	0.078	0.017 n.s
05 Tanzania						0.069	0.015 n.s	0.047	0.044	0.105	0.030	0.034	0.025
06 Zanzibar							0.081	0.074	0.084	0.108	0.096	0.110	0.033
07 Uganda 1								0.024 n.s	0.048	0.109	0.036	0.018 n.s	0.039
08 Uganda 2									0.082	0.123	0.048	0.055	0.045
09 DR Congo										0.113	0.047	0.056	0.028
10 Nigeria											0.079	0.115	0.083
11 Benin												0.015 n.s	0.051
12 Ghana L													0.074
13 Ghana M													

(first row) are found to encounter Sri Lanka genotypes in Africa, as Sri Lanka share only a small amount of genes with Africa. But several individuals of each African population have genotypes which can be assigned to Sri Lanka as seen in the higher probabilities values of the first column. We used the assignment approach to interpret the migration direction of *B. invadens* genotypes within Africa mainland and between Africa and Sri Lanka.

The average probability to encounter in the West (conservatively including Nigeria), genotypes from the East and vice versa, the average assignment probability value of East genotypes to West is 0.230±0.15, while the corresponding value from West to East is 0.113±0.10. Concerning the relation between Sri Lanka and African areas, eastern African genotypes are assigned to Sri Lanka with a relatively higher average probability (0.123±0.05) than western African genotypes (0.091± 0.07).

Genetic evidence for population bottlenecks

Although the method should be viewed with caution because the sample sizes for each population were less than 30 (Cornuet & Luikart 1996), in all samples no evidence of a recent population bottleneck based on a mode shift (i.e. paucity of rare alleles) was found. A significant heterozygote deficit was detected in Sri Lanka under the Stepwise Mutation Model, suggesting a recent population expansion in this island. Under the same model none of the invasive populations examined exhibit a significant genetic bottleneck in a two-tailed test for heterozygote excess and deficit (Wilcoxon rank test, Cornuet & Luikart, 1996).

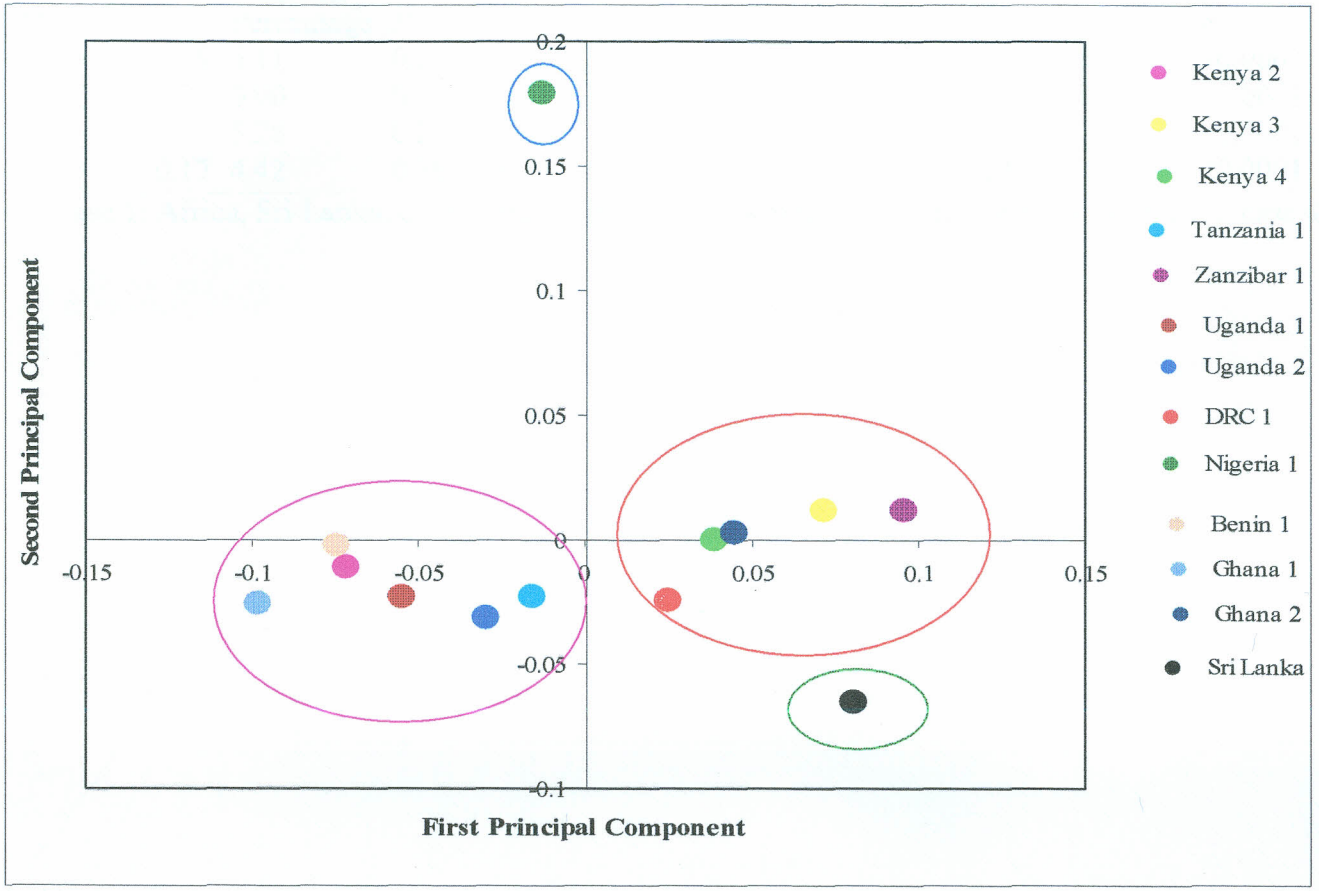


Fig. 11: Principal component graph showing the clusters separated by principal components 1 & 2

KENYATTA UNIVERSITY LIBRARY

Table 10 Analysis of molecular variance

	Among groups			Among populations within groups			Within populations		
Group	Va	Percentage	*P*	Vb	Percentage	*P*	Vc	Percentage	*P*
1	0.19	5.11	0.21	0.34	9.27	<0.0001	3.11	85.62	<0.0001
2	0.00	0.00	0.57	0.35	10.12	<0.0001	3.10	89.88	<0.0001
3	0.19	5.28	0.20	0.35	9.81	<0.0001	3.03	84.91	<0.0001
4	0.17	4.42	0.39	0.32	8.35	<0.0001	3.34	87.23	<0.0001

Case 1: Africa, Sri Lanka; case 2: East Africa, West Africa; case 3: East Africa, Sri Lanka; case 4: West Africa, Sri Lanka

Table 11: Average coefficient of ancestry obtained from a Structure run with K=4 for the 351 individuals of *B. invadens* from the 13 geographical regions. Co-ancestry higher than 10% of each population in a cluster is in bold

		Clusters (K)			
Area	**Population**	**1**	**2**	**3**	**4**
South Asia	Sri Lanka	0.030	0.015	0.014	**0.941**
East Africa	Kenya 2	0.025	**0.880**	0.086	0.010
	Kenya 3	**0.916**	0.016	0.049	0.019
	Kenya 4	**0.847**	0.034	0.090	0.029
	Tanzania	**0.257**	**0.637**	0.085	0.020
	Zanzibar	**0.700**	**0.171**	**0.104**	0.025
	Uganda 1	0.085	**0.864**	0.038	0.013
	Uganda 2	**0.198**	**0.700**	0.075	0.027
Central Africa	DR Congo	**0.791**	**0.104**	0.053	0.053
West Africa	Nigeria	0.030	0.014	**0.837**	**0.119**
	Benin	**0.134**	**0.618**	**0.197**	0.051
	Ghana 1	0.026	**0.930**	0.034	0.010
	Ghana 2	**0.711**	**0.161**	0.092	0.035

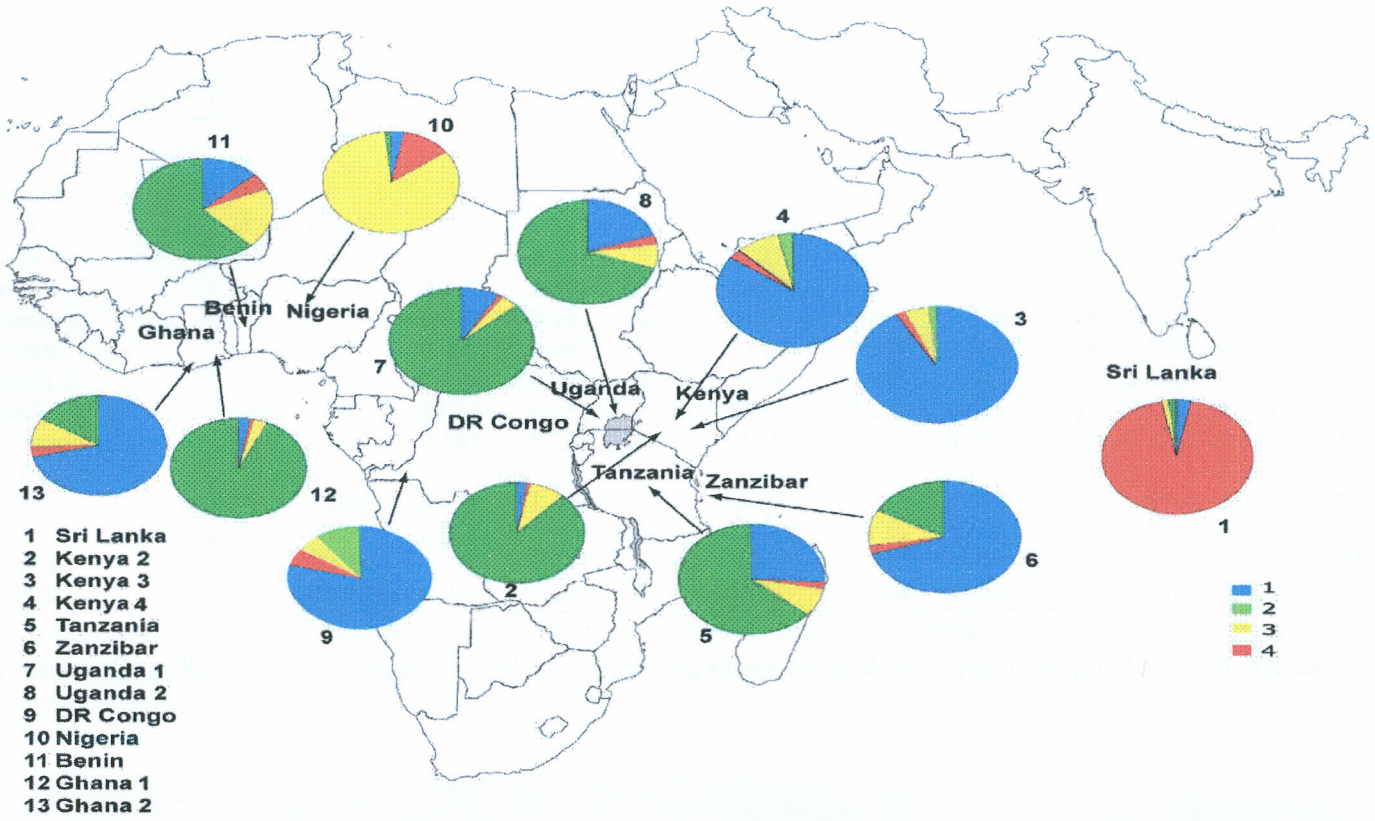


Fig 12 Geographical representation of the clustering outcomes for the 13 samples of *Bactrocera invadens*. The 4 colours represent the co ancestry distribution of the 351 individuals in 4 hypothetical clusters

Table 12: Mean assignment rate into (rows) and from (columns) each population as calculated by GENECLASS 2 (Piry *et al.*, 2004)

Assigned individuals	1	2	3	4	5	6	7	8	9	10	11	12	13
South Asia													
01 Sri Lanka	**0.401**	0.000	0.001	0.005	0.002	0.015	0.000	0.003	0.006	0.020	0.008	0.000	0.012
East Africa													
02 Kenya 2	0.087	**0.438**	0.048	0.117	0.402	0.256	0.411	0.396	0.162	*0.173*	*0.599*	*0.501*	*0.242*
03 Kenya 3	0.060	0.008	**0.507**	0.141	0.028	0.215	0.007	0.015	0.229	*0.078*	*0.071*	*0.014*	*0.130*
04 Kenya 4	0.169	0.037	0.187	**0.491**	0.210	0.390	0.091	0.209	0.357	*0.150*	*0.252*	*0.060*	*0.436*
05 Tanzania	0.188	0.219	0.083	0.191	**0.396**	0.293	0.313	0.308	0.228	*0.164*	*0.514*	*0.290*	*0.348*
06 Zanzibar	0.085	0.020	0.011	0.081	0.081	**0.454**	0.069	0.090	0.044	*0.079*	*0.090*	*0.051*	*0.223*
07 Uganda 1	0.080	0.174	0.019	0.099	0.339	0.247	**0.405**	0.403	0.184	*0.095*	*0.427*	*0.355*	*0.262*
08 Uganda 2	0.140	0.092	0.014	0.085	0.231	0.275	0.298	**0.501**	0.102	*0.116*	*0.354*	*0.216*	*0.285*
09 DR Congo	0.178	0.019	0.163	0.238	0.118	0.166	0.084	0.109	**0.488**	*0.120*	*0.257*	*0.080*	*0.315*
West Africa													
10 Nigeria	0.023	*0.004*	*0.010*	*0.013*	*0.010*	*0.031*	*0.001*	*0.005*	*0.015*	**0.490**	0.046	0.002	0.038
11 Benin	0.103	*0.087*	*0.028*	*0.074*	*0.127*	*0.103*	*0.116*	*0.133*	*0.154*	0.103	**0.408**	0.192	0.119
12 Ghana 1	0.056	*0.199*	*0.013*	*0.041*	*0.244*	*0.115*	*0.342*	*0.264*	*0.177*	0.089	0.561	**0.450**	0.133
13 Ghana 2	0.184	*0.027*	*0.066*	*0.230*	*0.142*	*0.351*	*0.086*	*0.153*	*0.249*	0.146	0.165	0.059	**0.421**

Values in bold indicate the proportions of individuals assigned to the source population. In italics are indicated the proportion of individuals from East Africa assigned to West Africa and those from West Africa assigned to East Africa.

Phylogenetic tree

Phylogenetic tree drawn by Phylip package using Cavalli Sforza chord distance gave two clusters as earlier illustrated by the Genalex program (Fig. 11). A cluster consisting of Benin, Ghana 1, Kenya 2, Tanzania 1, Uganda 1 and Uganda 2 and the other cluster consisting of DRC, Ghana 2, Kenya 3, Kenya 4 and Zanzibar 1. The Sri Lankan and Nigerian populations clustered on separate branches (Fig. 13).

5.5 Discussion

This pilot study attempts to infer from genetic data the dynamic aspects of the African invasion of the fruit fly pest, *B. invadens*. It has three principal goals: a) to confirm the Asian origin of this pest, b) to analyze the redistribution of genetic diversity within the invaded Africa mainland and c) to identify the route of the invasion in Africa. Few years (from 0 to 3) separate the first historical records of this species in Africa from the sampling years of this study, which covers a great part of the recent invasion area ranging from Kenya in the East to Ghana in the West. The first detection of this fly in a particular country doesn't necessarily mean that it was introduced at that time or in that country, as it could have been around for a much longer time at undetectable levels. However from the sampling sites the historical records are indicative of early stages of invasion because previous intensive fly trappings did not produce any evidence for the presence of this fly before 2000 (Lux *et al.*, 2003; Insect pest control newsletter 65, 2005; De Meyer *et al.*, 2009). Therefore, this study presents the first phases of the expansion events of the pest in Africa.

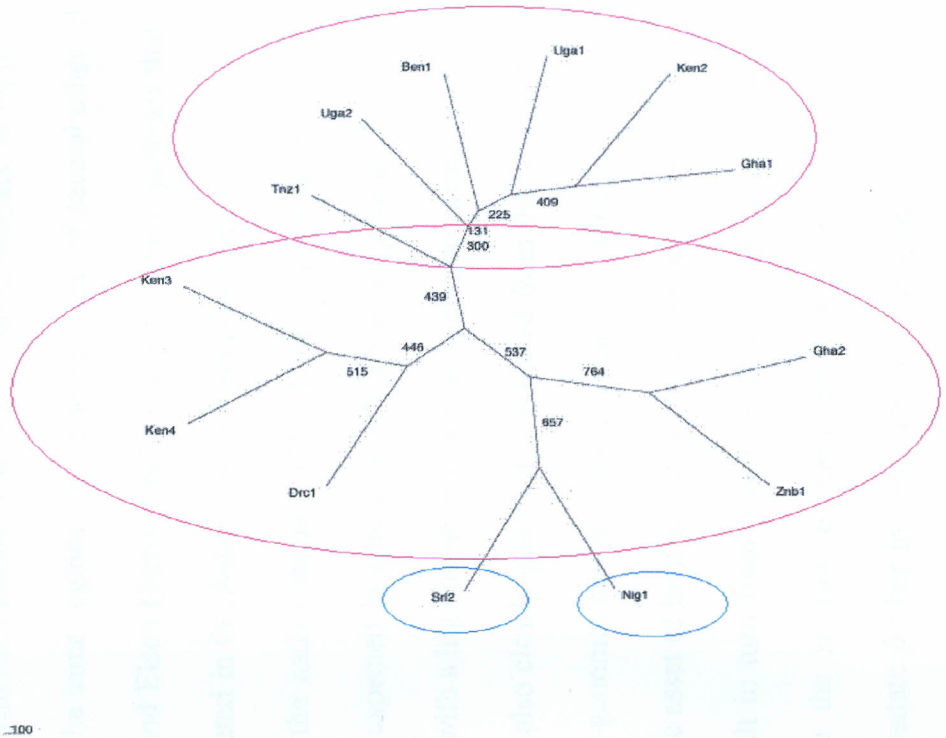


Fig. 13: Cavalli-Sforza distance rooted tree produced by the Phylip package

Correlative ecological models niche (ENMs), based on associations between known occurrence records of *B. invadens* and a set of environmental predictor variables, have recently been developed to define the ecological niche and the potential global distribution of this pest (De Meyer *et al.*, 2009). The results suggest that the species prefers hot and humid environments and the models clearly showed Sri Lanka and Southern India, with their humid endemic rainforest habitats, as highly suitable areas for this pest. Since the same regions are the source areas of several other sibling *Bactrocera* species (White and Elson-Harris, 1992), it supports the hypothesis that *B. invadens* may also have originated in the Asian subcontinent (Drew *et al.*, 2008). The genetic data cast no doubt about the native essence of Sri Lanka, as the sample is characterized by all genetic features expected in a large population in a native area, such as high number of alleles, coupled with a high number of private alleles occurring at high frequency. The Sri Lanka sample is also clearly genetically separated from the African flies and only a small percentage of its genomes can be found in Africa. It cannot be excluded that the peculiar Sri Lanka genetic asset of highest number of alleles per locus and high number of private alleles may result in turn from genetic isolation from the other southern Indian native areas. Therefore the relation between the native range and the origin of African infestation necessitate further investigation, including more samplings across the entire putative native range of the species.

The Equatorial rain forest belt and the East coastal regions of Africa, according to the ENM niche models (De Meyer *et al.*, 2009), are highly suitable for *B. invadens*. Indeed West Africa, Central Africa and part of East Africa are highly invaded areas and the

genetic data suggest the presence of populations with relatively high levels of genetic diversity associated with limited geographic structure. Furthermore, although the invasion is a relatively recent event in Africa, no genetic footprints of bottlenecks are present, but populations appear large enough to maintain, especially in the West a relatively high number of low frequency alleles. All these genetic features account for processes of rapid population growth and expansion. Indeed *B. invadens*, which is characterized by a high intrinsic rate of increase, may achieve enormous population size: the population is estimated to double in 6 days during a generation span of about 31 days (Ekesi *et al.,* 2006; 2009). Moreover the strong flying ability associated with very high polyphagy, involving both orchards and wild hosts, contribute to the dispersive power of these flies. Therefore this insect, in favorable environment, can be present all year-around, as it can switch to different suitable hosts when there is a bottleneck in carrying capacity. In the new African environment, all these features and the absence of natural enemies explain how this pest has adapted so quickly in the introduced range, also displacing native species of fruit flies (Ekesi *et al.,* 2009).

Genetic variability increases when moving from East to West. In the East, few months after the first discovery of *B. invadens* in the coastal regions of Kenya, a more thorough survey, showed that *B.invadens* was also present in western Kenya close to Uganda border (at Nguruman in the Rift Valley Province), as well as in Uganda and Tanzania. In Nguruman just one year after its detection in the country, *B. invadens* became the dominant fruit fly pest; the same rapid dominancy was observed in Morogoro region in Tanzania (Ekesi *et al.,* 2009; Mwatawala *et al.,* 2009). Two main reasons may account

for the slight lower level of variability of *B. invadens* in the East. One could be linked to the ecological niche as the insect appears to be adapted to a "hot-humid climate and lowland residence". In the East variable geographical and climatic regions with several arid areas are only partly indicated as suitable for *B. invadens* (De Meyer *et al.*, 2009). In addition to the above, there are also two major physical barriers to movement and survival of insect species across the East African plane. The volcanic and rifting activity that began 15-12 million years ago and continuing intermittently up to near present, created the Gregory Rift Valley, running approximately north-south through the most of the region (Griffiths 1993). The xeric floor of the Gregory Rift Valley that cut across most of the East African countries also acts as a formidable barrier for dispersal and also contains few, if any, plant species likely to be suitable hosts of pest tephritids. In addition, an extensive area of inhospitable habitat designated as "dry scrub with trees" (Greenway 1973) again runs approximately north-south that is unsuitable for tephritid survival.

By contrast, West Africa is a region of growing economic importance for fruit production and export, especially for mango. In fact this region plays a major role in local, national, regional, and international fruit market (Vayssieres *et al.*, 2009). During the 2005 and 2006 seasons, *B.invadens* was found to be the most important species in Benin and Ghana in terms of fly abundance and fruit damage. Given the lack of appropriate and systematic control methods and the invasive power of this pest, re-infestation of orchards from the widespread reservoirs of wild hosts, is a frequent and regular occurrence, with little fluctuations all year around. There is a supposition that in West Africa *B. invadens* was

already present since 2003 and probably developed in wild environment. In fact Mont Kouffe, the collection place of the highly variable Benin sample, is a protected area far from mango and other cultivated crops, with a vegetation composed of savannah, woodland and patches of dry forests. Wild places such as this constitute a reservoir of genetic variation: contemporary gene flow between highly cultivated areas and wild hosts contribute to the genetic diversity and probably to the creation of the mixed lineages observed in West Africa. A peculiar genetic asset is that of Nigeria population. The sample was collected in Kaduna region, one of the few agro-ecological region in Nigerian suitable for *B.invadens*: here, diverse small-holders cultivations of sweet orange, guava and mango, abundance of the umbrella tree *Terminalia* catappa (the most important alternative host of *B. invadens*) and lack of sanitation practices probably resulted in the settlement of large populations of the insect in the period from 2003 to 2006 (Umeh *et al.*, 2008; S. Ekesi, unpublished data).

It is clear, from the STRUCTURE analysis, that the population subdivision in Africa is not on geographical basis, as the Eastern African flies are not separated from Western flies. Instead the three African clusters, represented with the colors Blue, green, and yellow in Figure 12, can be interpreted as different invasion burst or outbreaks. Two main bursts originated in the East, probably from Kenya, as it is supported by the high proportional ancestry values of the three Kenyan samples to the Blue and green clusters. International trade in agricultural produce and travels between Eastern African and Asia countries which is on the increase may have resulted in inadvertent introduction. The yellow burst would originate from Nigeria, in the West, and appears more localized. It

could have resulted from independent introduction event(s) (perhaps through same process as above in addition to international passenger traffic along with shorter transit time), followed by little gene flow towards Benin and other western and eastern regions. Considering that Nigeria has certain degree of co-ancestry with Sri Lanka, we can hypothesize that this outbreak arose from an independent introduction event from an unsampled native source. However an alternative hypothesis is plausible: that Nigeria sample represents the outcome of a bottleneck (from Benin?), and that the high number of its rare alleles is a consequence of a relatively long period (2003-2006) of genetic isolation.

The genetic data demonstrate that the invasion and dispersal pattern of *B. invadens* in Africa was rapid and apparently chaotic, with possible multiple introductions as the hypothetical outbreaks would suggest. Looking at Table 12, the overall low 'self-assignment' values in the diagonal show that a maximum of 50% of the genomes in each population can be considered resident. In presence of such indication of an ongoing contemporary gene flow among invaded areas, and of post introduction hybridizations, it may appear difficult to recognize the route of the invasion. Nevertheless the two methods based on individual assignment, provided by STRUCTURE and GenClass2, allowed to infer the main pathway of dispersal. Two results are indicative: a) the higher or equal rate of co-ancestry of eastern with respect to the western flies in two African clusters; b) the major average assignment probability of eastern flies to the west than vice-versa. These two results support the fact that invasion started in East Africa. Regions of coastal East Africa, where *B. invadens* was firstly found, such as Kenya and /or Tanzania, on the basis

KENYATTA UNIVERSITY LIBRARY

of their high values of ancestry, consistently are the places from which the African invasion probably started. Another result is the major, although low, average assignment probability of East Africa flies to Sri Lanka than West Africa to Sri Lanka. Therefore genetic data are consistent with the supposition that East Africa was the port of entry of this fly.

KENYATTA UNIVERSITY LIBRARY

CHAPTER SIX

Identity of *Bactrocera invadens* in comparison with other *Bactrocera* species using DNA barcoding

6.1 Introduction

DNA barcoding is a molecular technique that involves retrieval of a standard region of mitochondrial gene, Cytochrome *c* oxidase 1 (COI) at its 5' end containing ≈ 650 base pairs gene to act as a 'barcode' to identify and delineate all animal life. The obtained sequence is specific for a particular species and therefore used as a 'tag' or barcode in identification of organisms in the same species. The two main goals of DNA barcoding are to assign unknown specimen to species and to enhance discovery of new species and facilitate identification, particularly in cryptic and other organisms with complex morphology (Hebert *et al.,* 2003a, b). This mitochondrial gene region is easily recovered and it provides good resolution, as evidenced by the fact that deep sequence divergences were the rule between 13,000 closely related pairs of animal species (Hebert *et al.,* 2003b).

Species identification through barcoding is usually achieved by the retrieval of a short DNA sequence the 'barcode' from a standard part of the genome (i.e. a specific gene region) from the specimen under investigation. The barcode sequence from each unknown specimen is then compared with a library of reference barcode sequences derived from individuals of known identity. A specimen is identified if its sequence closely matches one in the barcode library. Otherwise, the new record can lead to a novel barcode sequence for a given species (i.e. a new haplotype or geographical variant), or it can suggest the existence of a newly encountered species (Hajibabaei *et al.,* 2007). Various gene regions have been employed for species-level biosystematics;

however, DNA barcoding advocates the adoption of a 'global standard', and a 650-base fragment of the 5' end of the mitochondrial gene cytochrome c oxidase I (COI, cox1) (Hebert *et al.*, 2003a, b) has gained designation as the barcode region for animals.

Several studies have demonstrated the effectiveness of DNA barcoding in different animal groups (Hebert *et al.*, 2003a, b; Hebert *et al.*, 2004; Hajibabaei *et al.*, 2006; Smith *et al.*, 2006). These projects have shown that >95% of species possess unique COI barcode sequences; thus species-level identifications are regularly attained (Hajibabaei *et al.*, 2007). The earliest barcode studies received some criticism, mainly owing to their limited taxonomic and geographical breadth (Moritz and Cicero, 2004); however, more recent studies have addressed these issues by targeting species-rich groups (i.e. those containing many closely related species) in tropical settings (Hajibabaei *et al.*, 2006), and by comprehensive analyses of all the species in a given taxonomic assemblage (Meyer and Paulay, 2005). Momentum has further been aided by establishment of the Consortium for the Barcode of Life (CBOL, http://barcoding.si.edu) – an international alliance of research organizations that support the development of DNA barcoding as an international standard for species identification (Marshall, 2005) – and by development of the Barcode of Life Data Systems (http://www.barcodinglife.org) – a global online data management system for DNA barcodes (Ratnasingham. and Hebert, 2007). Barcoding projects typically involve gathering specimens of a given taxonomic group (identified by conventional taxonomic methods such as morphology; see below), cataloging them together with collateral data such as photographs and locality information, and assembling the barcode library (i.e. a 650-base segment of the COI gene) (Hajibabaei *et al.*, 2005).

The analysis of DNA barcoding data is usually performed by a clustering method, such as distance-based neighbor-joining (NJ) (Saitou & Nei 1987), and by evaluating genetic distances within and between species.

The amplification of the 650 bp region utilizes universal primers that are designed to amplify CO1 gene across broad taxa. Initially, the 5' end of COI was chosen as the focal region because it is flanked by two ''universal'' primers that work for a range of metazoans (Folmer *et al.,* 1994). The need to use widely applicable primers is understandable, but examination of the DNA barcoding literature reveals that the majority of projects actually rely on taxon specific primers, rather than universal primers, in order to optimize PCR performance (Hebert *et al.,* 2004; Penton *et al.,* 2004; Barrett and Hebert, 2005; Hebert *et al.,* 2005), particularly with degraded material (Lambert *et al.,* 2005). The DNA barcoding was used for identification of *B. invadens* by comparing it with other *Bactrocera* species such as *B. correcta, B. cucurbitae, B. dorsalis s.s, B. kandiensis, B. oleae B. paraverbascifoliae,* and *B. zonata.*

6.2 Materials and methods

6.2.1 Samples for the DNA barcoding

Alcohol preserved samples of *Bactrocera invadens* from Kenya (Nguruman (Ng)), Uganda (Ug), Nigeria (Zaria (Zr)) and Sri Lanka (Ranbukpitiya (SL)). Other *Bactrocera* species included in the analysis are: *B. correcta* (India), *B. cucurbitae* (Kenya), *B. dorsalis s.s* (Hawaii), *B. kandiensis* (Sri Lnka) *B. oleae* (Kenya), *B. paraverbascifoliae* (Sri Lanka) and *B. zonata* (Mauritius). Table 13 shows the samples and their collection details. All the insects used for analysis were photographed laterally, dorsally and ventrally and labelled appropriately.

6.2.2 DNA extraction

DNA from individual insects was extracted using the Qiagen DNeasy® Blood and Tissue Kit as per manufacturer's instructions (Qiagen). The DNA quality was checked by a 1% agarose gel. The samples were then stored at -20°C until use.

6.2.3 PCR amplification of the extracted DNA

DNA extracted from the insects was used as template for PCR amplifications. PCR was carried out using universal primers, Forward primer (LCO1490) 5'-GGTCAACAAATCATAAAGATATTGG-3' and reverse primer (HCO2198) 5'-TAAACTTCAGGGTGACCAAAAAATCA-3' (Folmer *et al.*, 1994). The PCR amplification were carried out in a volume of 20 µl containing 1x reaction Buffer, 200 µm of dNTP mix, 0.4 pmol/µl of each primer, 2.5 mM, $MgCl_2$, 1 units *Taq* DNA polymerase (Genescript) and 1 ng DNA template. Standard cycling conditions of 5 min at 94 °C, then 35 cycles of 30 s at 94 °C, 1 min at annealing temperature of 45 °C

Table 13: Populations used for DNA Barcoding analysis, locations and year of collection

Region/Country	Sample name	Sample site	Date	Coordinates
Bactrocera invadens				
Africa				
1. Kenya	Ke2	Nguruman	August 2008	01°48′32 S, 036°03′35 E
2. Uganda	Ug1	Kawanda	September 2007	00°49′52 S, 031°55′05" E
3. Nigera	Nig1	Zaria	November 2005	09°45′ 03 N, 080°23′ 00 E
Asia				
1. Sri Lanka	SL2	Ranbukpitiya	March 2007	11°06′ N, 07°42′ E
Bactocera correcta	Bcor	Sri Lanka-Anuradhapura	October 2007	08° 21' 0" N, 080° 23' 1" E
Bactrocera cucurbitae	Bcu	Kenya-Nairobi	June 2006	01° 13' 952 S, 036° 51' 314 E
Bactrocera dorsalis s.s	Bd	Hawaii	March 2007	Laboratory reared
Bactrocera kandiensis	BK	Sri Lanka-Kandy	July 2005	07°16′ 753 N, 80°35′ 731 E
Bactrocera oleae	Boleae	Kenya-Burguret forest	July 2008	00°06′ 720 S, 37° 02′ 342 E
Bactrocera paraverbascifoliae	Bp	Si Lanka-Ragala	October 2007	07° 16' 0 N, 80° 13' 60E
Bactrocera zonata	Bz	Mauritius	December 2008	Laboratory reared

and 1 min at 72 °C, followed by a final elongation step of 5 min at 72 °C were used. The PCR amplification products were by 1.5% agarose gel. PCR products were then purified using the Sodium acetate method.

6.3 Data analysis

Sequences were assembled and edited using Chromas version 2.13 (Copyright© 1998-2001 Technelysium Pty ltd, Queensland, Australia), and aligned in ClustalX version 1.81 (Jeanmougin *et al.,* 1998). Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 4 (Tamura *et al.,* 2007). Two trees were constructed; one by the Neighbour-joining option with bootstrapping and using the Kimura 2 distance matrix (Kimura, 1980) and the second tree by the Maximum parsimony option with bootstrapping. A table of between species distances was also constructed by the program MEGA 4.

6.4 Results

A total of 125 samples were extracted from DNA and amplified by the forward primer LCO1490 and reverse primer HCO2198. The PCR primers were optimised using the *Bactrocera invadens* DNA. The primers worked best at an annealing temperature of 45 °C (Fig. 18). The PCR products were then purified by Sodium acetate precipitation (Fig. 19) and sequenced. Sequencing was done with both the forward and reverse primers. PCR products were easily produced and aligned as no insertion, deletions or stop codons were observed. Additionally, no visualised PCR product contained double bands. These observations support the conclusion that the sequences analysed were mitochondrial DNA and not nuclear pseudogenes (Bensasson *et al.,* 2001). Sequences obtained were first edited by Chromas version 2.13 and then aligned by

ClustalX version 1.81. This provided an input file for the program Mega 4 from which a phylogenetic tree and distance table were generated. The evolutionary history was inferred using the Maximum Parsimony method (Eck & Dayhoff 1996). Tree number 1 out of 168 most parsimonious trees (length = 356) is shown. The consistency index is 0.661891, the retention index is 0.970448, and the composite index is 0.648783 for all sites and parsimony-informative sites 0.642331. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein, 1985). The MP tree was obtained using the Close-Neighbour-Interchange algorithm with search level 3 (Felsenstein, 1985; Nei & Kumar, 2000) in which the initial trees were obtained with the random addition of sequences (10 replicates). All positions containing gaps and missing data were eliminated from the dataset (Complete Deletion option). There were a total of 658 positions in the final dataset, out of which 181 were parsimony informative. The tree separated the *Bactrocera* species in clusters except for the *Bactrocera invadens* from Zaria (Nigeria) and Sri Lanka that clustered with the *B. kandiensis* (Fig. 16, also see appendix VI). The *B. invadens* from Kenya, Uganda, and some individual from the Nigeria and Sri Lankan populations clustered together. The number of base substitutions per site from averaging over all sequence pairs between groups is shown in Table 13. All results are based on the pairwise analysis of 125 sequences. Analyses were conducted using the Kimura 2-parameter method in MEGA4 (Tamura *et al.*, 2007). Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). There were a total of 658 positions in the final dataset. The *B. oleae* and *B. cucurbitae* samples generally had the highest values of the genetic distance, an indication that these two samples are genetically distanced from the other *Bactrocera*

species. The lowest distance (0.04) is between *B. invadens* Kenya and Uganda, confirming that these two samples are very close genetically. Likewise, these two samples share a close relationship with the *B. paraverbascifoliae*. The highest mean distance within group is 0.04 in *B. invadens* Sri Lanka and 0.027 in *B. invadens* Zaria. The rest of the samples had a mean distance of between 0 – 0.09. The Tajima test statistic (Tajima 1989) was also estimated using MEGA4 (Table 14). All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). The sequences had a total of 125 sites, 188 segregating sites, at a Tajima test statistic of D = 1.615185.

Maximum composite likelihood estimate of the pattern of nucleotide substitution was also analysed by the program MEGA4 (Table 15). Each entry shows the probability of substitution from one base (row) to another base (column) instantaneously.

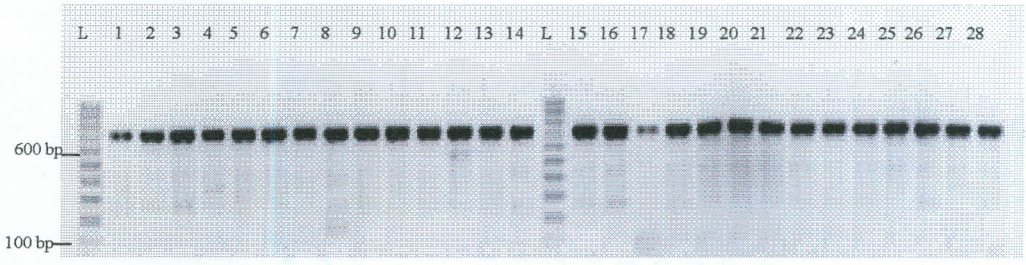


Fig. 14: A representative gel of PCR products amplified by the LCO1490 and HCO2198 at 45 °C. L - 100 bp ladder, 1 – 28 are samples analysed.

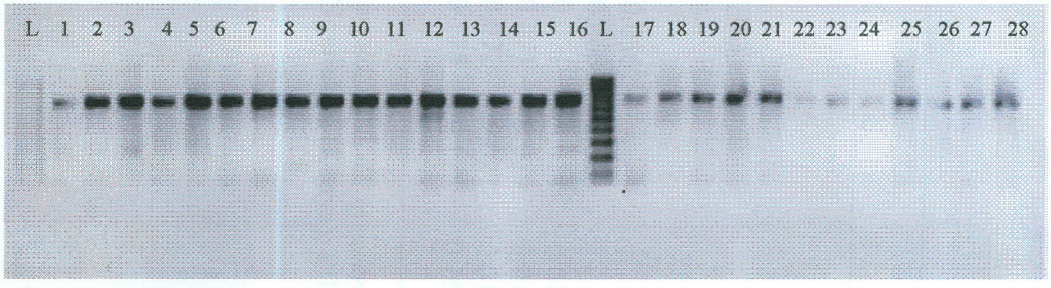


Fig. 15: A representative gel of Purified PCR products for sequencing. L – 100 bp ladder. 1 – 28 are samples purified.

KENYATTA UNIVERSITY LIBRARY

Table 14: Means of Genetic distances between populations generated by Mega 4 program (Tamura *et al.*, 2007)

	1	2	3	4	5	6	7	8	9	10	11
1. Bcu											
2. Bcor	0.183										
3. Bz	0.182	0.076									
4. BiSL	0.182	0.098	0.103								
5. BK	0.187	0.099	0.100	0.031							
6. BiZr	0.175	0.098	0.103	0.044	0.056	-					
7. Bdorsalis	0.166	0.091	0.101	0.046	0.060	0.023	-				
8. BiNg	0.171	0.099	0.105	0.049	0.071	0.018	0.014	-			
9. BiUg	0.172	0.099	0.106	0.050	0.071	0.019	0.014	0.004	-		
10. Bp	0.165	0.095	0.107	0.051	0.071	0.021	0.015	0.007	0.007	-	
11. Boleae	0.194	0.176	0.183	0.176	0.175	0.175	0.169	0.176	0.177	0.171	-

Bcu- *Bactrocera cucurbitae*, Bcor-*B. correcta*, Bz-*B. zonata*, BiSL-*B. invadens* Sri Lanka, BK-*B. kandiensis*, BiZr-*B. invadens* Zaria (Nigeria), Bdorsalis-*B. dorsalis s.s*, BiNg-*B. invadens* Nguruman (Kenya), BiUg-*B. invadens* Uganda, Bp-*B. paraverbascifoliae* and Boleae-*B. oleae.*

Table 15: Results from Tajima's Neutrality Test for 125 sequences

M	S	p_s	Θ	Π	D
125	188	0.285714	0.052895	0.078923	1.615185

The abbreviations used are as follows: m = number of sites, S = Number of segregating sites, p_s = S/m, $\Theta = p_s/a_1$, and Π = nucleotide diversity. D is the Tajima test statistic. a_1 is the gamma parameter. According to the gamma distribution, the substitution rate often varies from site to site within a sequence. The shape of this distribution is determined by a value known as the gamma parameter, which is also known as the shape parameter.

Table 16: Maximum Composite Likelihood Estimate of the Pattern of Nucleotide Substitution

	A	T	C	G
A	-	*2.9*	*1.59*	**13.28**
T	*2.39*	-	**16.82**	*1.4*
C	*2.39*	**30.66**	-	*1.4*
G	**22.7**	*2.9*	*1.59*	-

A, T, C and G are nucleotides (Adenine, Thiamine, Cytocine and Guanine). In bold, are transitional substitution rates of the nucleotides while in italics are transversional rates.

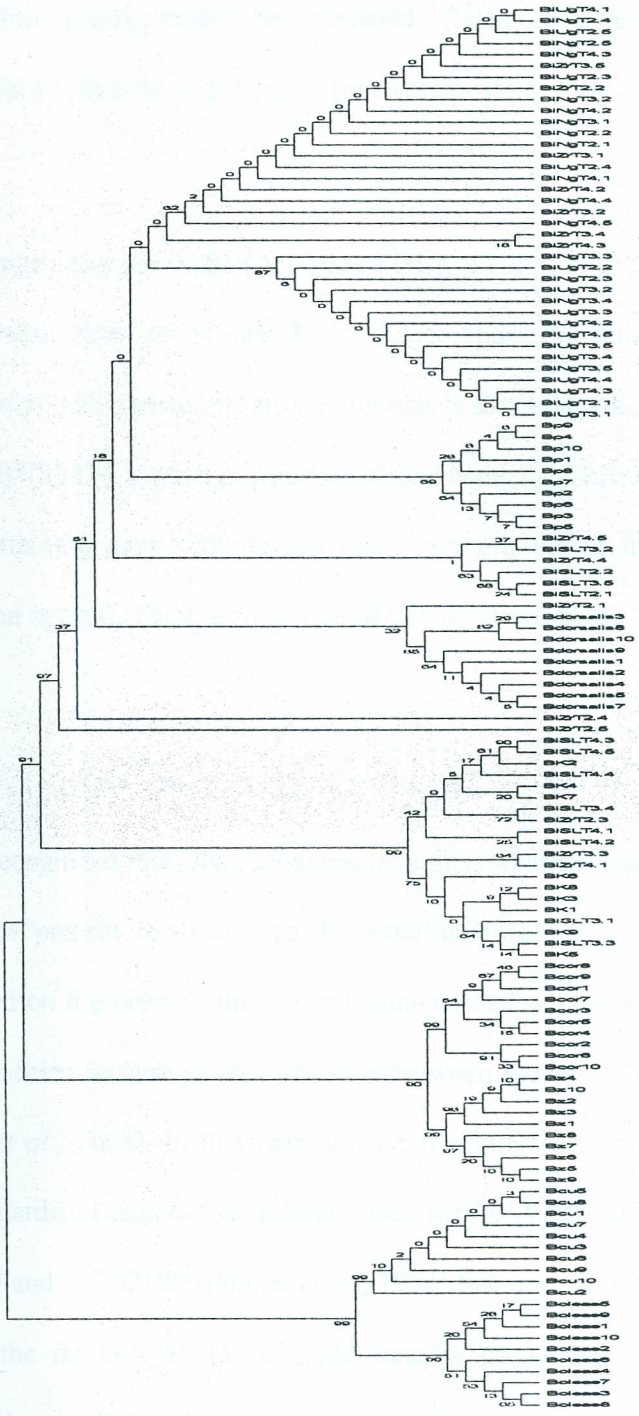


Fig. 16 Maximum Parsimony phylogenetic tree generated by MEGA4 (Tamura *et al.*, 2007) (Appendix VI).

Only entries within a row should be compared. Rates of different transitional substitutions are shown in bold and those of transversional substitutions are shown in italics.

The nucleotide frequencies are 0.288 (A), 0.351 (T/U), 0.192 (C), and 0.169 (G). The transition/transversion rate ratios are k_1 = 9.515 (purines) and k_2 = 10.572 (pyrimidines). The overall transition/transversion bias is R = 3.88, where R = [A*G*k_1 + T*C*k_2]/[(A+G)*(T+C)]. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated from the dataset (Complete-deletion option). There were a total of 658 positions in the final dataset.

6.5 Discussion

It has long been recognized that DNA sequence diversity, whether assessed directly or indirectly through protein analysis, can be used to discriminate species. DNA barcoding is based on the premise that a short standardized sequence can distinguish individuals of a species because genetic variation between species exceeds that within species (Hebert *et al.*, 2003). In this case, universal primers are employed that can amplify this standardized region. The primers used for DNA barcoding in this study are the LCO1490 and HCO2198 (Folmer *et al.*, 1994) that successfully amplified the CO1 region of the mtDNA at 45 °C. Additionally, no visualised PCR product contained double bands. These observations support the conclusion that the sequences analysed were mitochondrial DNA and not nuclear pseudogenes (Bensasson *et al.*, 2001). Moreover, the alignment of COI sequences was straightforward, as indels were uncommon, reinforcing the results of earlier work showing the rarity of indels in this

gene (Mardulyn & Whitfield 1999). Aside from their ease of acquisition and alignment, the COI sequences possessed, as expected, a high level of diversity.

The interspecific variability was higher than the intrapecific variability as expected for the populations analysed. The maximum parsimony tree showed the different species clustering together except for the *Bactrocera kandiensis* clustering with some of the individuals from *B. invadens* from Sri Lanka and Nigeria. While barcode libraries have similarities to molecular phylogenetic data (both are sequence information from assemblages of species), DNA barcodes do not usually have sufficient phylogenetic signal to resolve evolutionary relationships, especially at deeper levels (Hajibabaei *et al.,* 2006). Although its role in identifying specimens to a species level is an important aid for taxonomic workflow barcoding is no replacement for comprehensive taxonomic analysis (Hajibabaei *et al.,* 2007). Although barcode sequences have been analyzed mainly by using phylogenetic tree reconstruction methods such as NJ, these barcode-based trees should not be interpreted as phylogenetic trees (Hajibabaei *et al.,* 2006).

Barcoding using the CO1 gene enabled the interpretation of the relationship between the new pest to science, *Bactrocera invadens* and the other *Bactrocera* species. The smallest genetic distance detected (0.004) is between the *B. invadens* from Kenya and Uganda. The *B. invadens* from all the localities analysed seem to be in close genetic proximity to *B. dorsalis s.s*, *B. kandiensis* and *B. paraverbascifoliae*. This suggests the need for other molecular markers to be used in order to interpret the relationship between them. This is because the CO1 gene is just a small region of 650 bp.

CHAPTER SEVEN

GENERAL DISCUSSION, CONCLUSION AND RECOMMENDATIONS

7.1 General discussion

The economic impact of exotic species has led to an increased interest in studies of their ability to disperse, colonize and become established (Lodge 1993; Holway & Suarez 1999). The term ''biological invasion'' highlights the negative aspects of population migration represented by the impact on biodiversity, human economy and/or health. Biological invasions have recently become a subject of numerous studies in ecology, evolutionary biology, population biology and genetics (Saki *et al.,* 2001; Lee, 2002). The attention is centred on contemporary biological invasions and on the role of humans in accelerating the introduction rate of non-native species into new environments. Although invasions are old as life itself, there is no doubt that biological invasions will continue to increase in frequency as global trade increases. Yet we often have a poor understanding of the biological and genetic traits that make a species a good invader.

The phytophagous insects of the Tephritidae family, commonly referred to as ''true fruit flies'', offer different case histories of successful invasions, mainly human mediated. This family includes more than 5,000 species worldwide, approximately 1,400 of which develop in fleshy fruits. The four major genera of this family, *Ceratitis, Bactrocera, Anastrepha* and *Rhagoletis,* include important pest species (White and ElsonHarris, 1992). Each genus has a proper natural distribution: *Ceratitis* is an Afrotropical genus, *Bactrocera* is mainly confined to the Oriental and Australasian regions, *Anastrepha* to

South and Central America and the West Indies, while *Rhagoletis* has representatives in the Americas, Europe and temperate Asia.

In recent years, members of these genera have been reported outside their original range as the result of the growing fruit trade and tourist industry. The spread of these species is receiving considerable attention and is mobilizing substantial ecological and evolutionary genetic research (Aluja and Norrbom, 2000). The fruit flies differ in many aspects including their hosts, mating behaviour, generation time, survival in unfavourable periods, dispersal characteristics, etc. This diversity for traits directly related to fitness may imply different and substantially unpredictable invasion potentials for each species.

In March 2003 an invasive and devastating quarantine fruit fly pest of Asian origin was first detected in Kenya and was later described as *Bactrocera invadens* because of its rapid spread over large geographical areas. This pest has subsequently been reported from 30 other countries across the African continent including Angola, Benin, Burkina Faso, Cameroon, Central African Republic, Chad, Comoros, Congo, Congo Democratic Republic, Cote d'Ivoire, Equatorial Guinea, Ethiopia, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Kenya, Liberia, Mali, Mauritania, Mozambique, Niger, Nigeria, Senegal, Sierra Leone, Sudan, Tanzania, Togo, Uganda, Zambia and three regions in Asia including Bhutan, India and Sri Lanka. Due to the novelty status of *B. invadens* there had been little known about the morphometric and genetic diversity of this pest. As invasive species spread around the globe, it is becoming increasingly evident that a detailed knowledge of the biology, genetic structure and geographical variability, of a given

species is a prerequisite to planning strategies for quarantine, control or eradication (Roderick & Navajas 2003).

In this study morphometry using microscopy and screen measuring software of the right wing and hind tibia was employed to separate the *B. invadens* populations and compare it with other *Bactrocera* species. The *Bactrocera invadens* populations and the other *Bactrocera* species belonging to the *B. dorsalis* complex could not be separated by PCA. However, the first two principal components separated the *B. correcta*, *B. cucurbitae*, *B. oleae* and *B. zonata* into distinct groups. Likewise, the projection of the data on the first two canonical variate axes showed a similar pattern of separation. The *B. invadens* populations and the other *B. dorsalis* species complex clustered together while *B. correcta*, *B. cucurbitae*, *B. oleae* and *B. zonata* distinctly separated. When analysis was done considering only the Sri Lankan *B. invadens*, *B. dorsalis s.s* and *B. kandiensis*, the Sri Lankan *B. invadens* was closer to *B. kandiensis* (with $D^2 = 7.2$) than *B. dorsalis s.s* ($D^2 = 38.5$). The Sri Lankan *Bactrocera invadens* and *B. kandiensis* could be separated by neither the canonical variate nor the principal component analyses an indication that the *B. invadens* is closely related to the *B. kandiensis*. Morphometric methods are frequently used for identification of populations within species and it is often necessary to use more than one character to discriminate between groups. According to Richtsmeier *et al.* (2002), the first approach involves the application of quantitative methods to discover new information within the data and then looking for patterns within the data that suggest any underlying biological processes or effects. To analyse such large data sets, computer programmes of multivariate methods are needed (Sneath & Sokal 1973),

KENYATTA UNIVERSITY LIBRARY

and some of the commonest methods include (among others) principal components and canonical variates analyses. Projection of the populations on the first two principal axes showed a partial separation of the populations.

For population genetic structure and dynamic studies of the *B. invadens,* 11 polymorphic microsatellite markers were isolated and tested to assess their polymorphism in two wild populations: Sri Lanka and Democratic Republic of Congo. These loci amplified all the *B. invadens* colour forms uniformly. All loci were polymorphic across the two samples and the number of alleles per locus ranged from three to fifteen and eight loci had a polymorphic information content PIC > 0.5. The microsatellite repetition more frequently observed in *B. invadens* is the $(CA)_n$, which is also observed in the *B. cacuminata* (Song *et al.*, 2006), *B. dorsalis s.s* (Aketarawong *et al.*, 2006) and *B. oleae* (Augustinos *et al.*, 2002). All loci isolated were used for population genetic study. Microsatellite markers have been successfully applied to other invasive fruit fly species to infer evolutionary aspects underlying their invasive processes and to identify the routes of their colonization (Bonizzoni *et al.* 2001, 2004; Baliraine *et al.* 2004; Augustinos *et al.* 2005; Nardi *et al.* 2005; Akaterawong *et al.* 2007).

The population genetic structure of *B. invadens* was assessed using samples from across Africa and a sample from the putative aboriginal home of Sri Lanka. The analysis of microsatellite variability in the *B. invadens* samples collected across the actual species distribution range aimed to unravel the colonization structure and migration routes in and among different areas, and possibly to identify the source area of this invasive pest

KENYATTA UNIVERSITY LIBRARY

species. All the 13 geographical samples, including the Si Lankan sample, displayed a high level of variability. The presence of highest frequency of private alleles in the Sri Lankan sample, suggests that this sample comes from a large and well established population. The oldest population among the samples analysed, seem to be the Sri Lankan population. The genetic data cast no doubt about the native essence of Sri Lanka, as the sample is characterized by all genetic features expected in a large population in a native area, such as high number of alleles, coupled with a high number of private alleles occurring at high frequency. The Sri Lanka sample is also clearly genetically separated from the African flies and only a small percentage of its genomes can be found in Africa. Therefore the relation between the native range and the origin of African infestation necessitate further investigation, including more samplings across the entire putative native range of the species.

When STRUCTURE was employed to obtained genetic clustering of the *B.invadens* populations analysed using the Bayesian clustering method and determining the clusters by the method of Evanno *et al.* (2005), four hypothetical clusters were obtained showing that all the 351 individuals from the 13 populations could be categorized into four clusters. This clustering isolated the Sri Lankan population and Nigerian populations into separate clusters while the other African populations clustered into the other two clusters with no discrimination between the east and west African samples.

The phylogenetic tree drawn from the distance by Cavalli Sforza, also separated the populations into four clusters as shown in Fig. 13 isolating the Nigerian population from the other African populations and Sri Lankan.

Cross amplification of the *Bactrocera dorsalis s.s* primers previously developed by Aketarawong *et al.* (2006) on the *B. invadens*, was successful for the following primers: Bd1, Bd6, Bd9, Bd15, Bd19, Bd42, Bd43a, Bd76 and Bd99 out of the 20 loci isolated. Failure to cross amplify with the other primers suggests that the two are closely related but not the same species, as previously thought. This information is crucial for management strategies and proper prediction of the best control measure for control or eradication of the pest.

DNA barcoding is a molecular technique that involves retrieval of a standard region of mitochondrial gene, Cytochrome c oxidase 1 (CO1) at its 5' end containing ≈ 650 base pairs gene to act as a 'barcode' to identify and delineate all animal life. The obtained sequence is specific for a particular species and therefore used as a 'tag' or barcode in identification of organisms in the same species. The primers used for DNA barcoding in our study are the LCO1490 and HCO2198 (Folmer *et al.*, 1994) that successfully amplified the CO1 region of the mtDNA at 45 °C. The interspecific variability was higher than the intrapecific variability as expected for the populations analysed (Table 14). The maximum parsimony tree showed the different species clustering together except for the *Bactrocera kandiensis* clustering with some of the individuals from *B. invadens* individual from Sri Lanka and Nigeria.

The *B. invadens* from all the localities analysed seem to be in close genetic proximity to the *B. dorsalis s.s*, the *B. kandiensis* and the *B. paraverbascifoliae*. These results confirm the placement of the *Bactrocera invadens* as a separate species from the *B. dorsalis s.s.*

The rapid demand for tropical fruits and increased importance of global market is parallel to the increasing risks of invasion by both, known and undiagnosed species. Few years after its discovery in East Africa and its identification as a member of *Bactrocera* complex, *B. invadens* samples from all equatorial Africa showed a high level of genetic diversity associated with an evident absence of geographic structure. The association of these two aspects reveals that *B. invadens* was not an indigenous African species that remained undetected for long period of time. Instead this species is a recent and an aggressive invader. Data from the timing of historical records, that indicated this insect as a new entrant in a region, are in this case concordant with the chronology of the spread as documented in this study.

7.2 Conclusions

1. Morphometric analysis of the *Bactrocera invadens* showed that morphotypes can not be separated using the 14 veins of the right wing and the hind tibia of the insect i.e there was no significant differences between morphotypes using the multivariate analysis undertaken. The only differences are in the scutum coloration.
2. Morphometry also could not separate the different populations of the *B. invadens* from across Africa and Sri Lanka using the right wing and hind tibia.
3. Morphometric analysis separated the *B. invadens* from the *B. correcta, B. cucurbitae, B. oleae* and *B. zonata. Bactrocera invadens* clusters together with *B.*

dorsalis s.s and *B. kandiensis* and could not be separated from these two by neither the Principal component no the Canonical variate analyses.

4. When analysis was performed considering the *B. invadens* from Sri lanka, *B. dorsalis s.s* and *B. kandiensi. B. invadens* was closer to the *B. kandiensis* with the least mahalonobis squared distance.

5. Morphometric work was confirmed by the microsatellite analysis, that there were no differences in the morphotypes. Also the barcoding results confirmed the closeness of the *B. invadens* to the *B. kandiensis*.

6. Eleven polymorphic microsatellite markers were successfully isolated from the *Bactrocera invadens*. These are the first set of markers to be isolated from this species.

7. The eleven loci were applied successfully to infer the invasion history of the *Bactrocera invadens* in Africa and the presumed aboriginal home of Sri Lanka.

8. The Sri Lanka sample has the highest number of alleles and private alleles. It is therefore a well established population.

9. Genetic differentiation in the *B. invadens* populations quantified by pairwise FST values, are as shown in Table 9. Pairwise F_{ST} values were generally low, with F_{ST} ranging from 0.015 to 0.134. Low values of differentiation are observed among Eastern African populations. In the West Africa, Nigeria has the highest values of F_{ST} with all eastern African populations, but this population shows a certain degree of isolation also among the western populations.

10. Results of STRUCTURE analysis showed a maximum for $\Delta K = 4$, an indication that the 351 individuals in the 13 populations can be described within 4 hypothetical genetic subdivision defined as clusters.

11. In a principal component analysis, the first three principal coordinates accounted for 27.63%, 23.55% and 20.15% of the variance, respectively. The first two principal coordinate's revealed two distinct African clusters while Nigerian and Sri Lankan samples clustering on their own.

12. In all samples no evidence of a recent population bottleneck based on a mode shift (i.e. paucity of rare alleles) was found. A significant heterozygote deficit was detected in Sri Lanka under the Stepwise Mutation Model, suggesting a recent population expansion in this island.

12 The lack of strong regional structure among our samples was also evidenced by AMOVA (Table 10). Up to (not significant) 5% of genetic variance results partitioned among groups in the four tests performed, where the bulk of genetic variation was within groups and among individuals within populations.

13 When AMOVA was analysed without considering Sri Lanka, and only African samples are compared, the eastern and western African samples show no proper molecular variance.

14 Phylogenetic tree for the microsatellite markers drawn by Phylip package gave two African clusters with no discrimination between the East and West African samples. The Sri Lankan and Nigerian sample also clustered on separate branches.

13. Barcode analysis showed that the *B. invadens* from Nigeria and Sri Lanka are close to the *B. kandiensis*.

7.3 Recommendations

Further analysis to separate the species that are closely related when separated by the COI gene examples the *Bactrocera invadens*, *B. kandiensis* and *B. paraverbascifoliae* should be undertaken. Due to the clustering together of the species when considering the COI gene, effort should be taken to resolve the *dorsalis* complex using other markers like the Cyt b or the COI – COII intergenic region. More sampling of the *Bactrocera invadens*, *B. dorsalis* and *B. kandiensis* should be done and analysis performed using the microsatellite markers obtained in this study and also the COI gene.

This study has also indicated that there are no genetic differences when using the microsatellite markers obtained in this study and COI gene on the morphotypes. Also the closeness of *Bactrocera invadens* to the other *Bactrocera* species belonging to the dorsalis complex should be an indication that control strategies for instance the use of parasitoids, can be transferred from these other species to the *Bactrocera invadens*.

REFERENCES

Aketarawong N., Bonizzoni M., Malacrida A.R., Gasperi G., Tanaphum S. (2006). Seventeen novel microsatellite markers from an enriched library of the pest species *Bactrocera dorsalis sensu strictu*. Molecular Ecology Notes, 6: 1138-1140.

Aketarawong N., Bonizzoni M., Thanaphum S., Gomulski L.M., Gasperi G., Malacrida A.R., Gugliemino C.R. (2007). Inferences on the population structure and colonization process of the invasive oriental fruit fly, *Bactrocera dorsalis* (Hendel). *Molecular Ecology,* 16, 3522-3532.

Aluja M., Norrbom A.L. (2000). *Fruit Flies (Tephritidae): Phylogeny and Evolution of Behavior*. CRC Press, Boca Raton, Florida.

Armstrong K.F., Cameron C.M. & Frampton E.R. (1997a). Fruit fly (Diptera: Tephritidae) species identification: a rapid molecular diagnostic technique for quarantine application. *Bulletin of Entomological Research,* **87**: 111-118.

Augustinos A.A., Mamuris Z., Stratikopoulos E.E., D'Amelio S., Zacharopoulou A., Mathiopoulos K.D. (2005). Microsatellite analysis of olive fly populations in the Mediterranean indicates a westward expansion of the species. *Genetica*, **125**, 231-241.

Averill, A.L. & Prokopy, R.J. (1989). Host marking pheromones. In: *Fruit flies: their biology, natural enemies and control* (A. S. Robinson and G. Hooper, eds); Elsevier, Amsterdam. *pp*: 207-219.

Bachtrog D., Weiss S., Zangerl B., Brem G., Schlötterer C. (1999). Distribution of dinucleotide microsatellites in the *Drosophila melanogaster* genome. Molecular Biology and Evolution, 16: 601-610.

Baliraine F.N., Bonizzoni M., Osir E.O., Lux S.A., Mulaa F.J., Zheng L., Gomulski L.M., Gasperi G. and Malacrida A.R. (2002). Comparative analysis of microsatellite loci in four fruit fly species of the genus *Ceratitis* (Diptera: Tephritidae). *Bulletin of Entomological Research.*

Barrett R.D.H. & Hebert P.D.N. (2005). Identifying spiders through DNA barcodes. *Can. J. Zool.* **83**(3): 481–491.

Baruffi L., Damiani G., Guglielmino C.R., Bandi C., Malacrida A.R., Gasperi G. (1995). Polymorphism within and between populations of *Ceratitis capitata*: comparison between RAPD and multilocus enzyme electrophoresis data. *Heredity*, **74**, 425-437.

Bateman M.A. (1972). The ecology of fruit flies. *Annual Review of Entomology* **17**: 493-517.

Bensasson D., Zhang D., Hartl D.L., Hewitt G.M. (2001). Mitochondrial pseudogenes: Evolution's misplaced witnesses. *Trends Ecol Evol* **16**: 314–321.

Berry S.D., Fondong V.N., Christine R., Rogan D., Fauquet C.M. and **Brown J.K. (2004).** Molecular Evidence for five distinct *Bemisia tabaci* (Homoptera: Aleyrodidae) geographic haplotypes associated with cassava plants in sub-Saharan Africa. *Annals of the Entomological Society of America,* **97:** 852 – 859.

Billah M.K. (2004). Biosystematic Studies of *Psyttalia* species *(Hymenoptera: Braconidae):* Parasitoids of attacking fruit-infesting flies *(Diptera: Tephritidae) in Africa.* PhD Thesis. University of Ghana, Legon-Accra. 236 pp.

KENYATTA UNIVERSITY LIBRARY

Billah, M.K., Kimani-Njogu, S., Overholt, W.A., Wharton, R.A., Wilson, D.D. & Cobblah, M.A. (2005). The effect of host larvae on three Psyttalia species (Hymenoptera:Braconidae): parasitoids of fruit-infesting flies (Diptera: Tephritidae). *International Journal of Tropical Insect Science* **25(3)**, 168–175.

Bonizzoni M., Malacrida A.R., Guglielmino C.R., Gomulski L.M., Gasperi G. and Zheng L. (2000). Microsatellite polymorphism in the Mediterranean fruit fly *Ceratitis capitata*. *Insect Molecular Biology* **9**: 251-261.

Bonizzoni M., Zheng L., Guglielmino C.R., Haymer D.S., Gasperi G., Gomulski L.M. & Malacrida A.R. (2001). Microsatellite analysis of medfly bioinfestations in California. *Molecular Ecology* **10**: 2515-2524.

Bonizzoni M., Guglielmino C.R., Smallridge C.J., Gomulski L.M., Malacrida A.R., Gasperi G. (2004). On the origins of medfly invasion and expansion in Australia. *Molecular Ecology*, 13, 3845-3855.

Bowcock A.M., Ruiz-Linares A., Tomfohrde J., Minch E., Kidd J.R., Cavalli-Sforza L.L. (1994). High resolution of human evolutionary trees with polymorphic microsatellites. *Nature*, **368**, 455-457.

Bookstein F.L., Chernoff B., Elder R., Humphries J., Smith G. & Strauss R. (1985). Morphometrics in evolutionary biology. *The Academy of Natural Science of Philadelphia, Special Publication*, No. **15**: 277 pp.

Brookfield J.F.Y. (1996). A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology*, **5**, 453-455.

Brown W.M., George Jr. M. & Wilson A.C. (1979). Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences, USA,* **76:** 1967 – 1971.

Bruford M.W. & Wayne R.K. (1993). Microsatellite and their application to population genetic studies. *Curr. Opin. Genet. De.*, 3: 939-943.

Bubliy O.A., Loeschcke V., Imasheva A.G. (2001). Genetic variation of morphological traits in *Drosophila melanogaster* under poor nutrition: isofemale lines and offspring-parent regression. *Heredity*, **86**: 363–369.

Bubliy O.A. & Loeschcke V. (2002). Effect of low stressful temperature on genetic variation of five quantitative traits in *Drosophila melanogaster. Heredity* **89**, 70–75.

Caterino M.S., Cho S. & Sperling F.A.H. (2000). The current state of insect molecular systematics: a thriving tower of babel. *Annual Review of Entomology* **45:**1–54.

Caterino M.S. & Vogler A.P. (2002). The phylogeny of Histeroidea (Coleoptera: Staphyliniformia). *Cladistics*, **18:** 394 – 415.

Cavalli-Sforza L.L. & Edwards A.W. (1967). Phylogenetic analysis. Models and estimation procedures. *American Journal of Human Genetics*, **19**, 233–257.

Cervera M.T., Cabezas J.A., Simon B., Martinez-Zapater J.M., Beitia F. & Cenis J.L. (2000). Genetic relationships among biotypes of *Bemisia tabaci* (Hemiptera: Aleyrodidae) based on AFLP analysis. *Bulletin of Entomological Research,* **90:** 391 – 396.

Christenson L.D. & Foote R.H. (1960). Biology of fruit flies. *Annual Review of Entomology* **5**: 171-192.

Clarke A.R., Armstrong K.F., Carmichael A.E., Milne J.R., Raghu S., Roderick G.K. & Yeates D.K. (2005). Invasive phytophagous pests arising through a recent

evolutionary radiation: the *Bactrocera dorsalis* complex of fruit flies. *Annual Review of Entomology* **50**, 293-319.

Clary D.O. & Wolstenholme D.R. (1985). The mitochondrial DNA molecule of Drosophila yakuba: neucleotide sequence, gene organisation and genetic code. *Journal of Molecular Evolution,* **22:** 252 – 271.

Clausen C.P. (1978). Tephritidae (Trypetidae: Trupaneidae). Introduced Parasite and Predators of Arthropod Pest and Weeds: *A World Review* (C. P. Clausen, ed); U.S. Department of Agriculture Handbook No, **480**. *pp*: 320-335.

Cogan B.H. & Munro H.K. (1980). 40. FamilyTephritidae pp. 518-554. *In*: Crossskey, R.W. (Ed), Catalogue of the Diptera of the Afrotropical Region. *British Museum (Natural History), London.*

Copeland R.S., Wharton R.A., Quentin L. & De Meyer M. (2002). Indigenous hosts *Ceratitis capitata* (Diptera: Tephritidae) in Kenya. *Annals of Entomological Society of America*, **95**(6: 672-694).

Copeland R.S., White I.M., Okumu M., Machera P., Wharton R.A. (2004). Insects associated with fruits of the Oleaceae (Asteridae, Lamiales) in Kenya, with special reference to the Tephritidae (Diptera). *Bishop Museum Bulletin in Entomology*, 12, 135–164.

Cornuet J.M. & Luikart G. (1996). Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144**, 2001-2014.

De Meyer M. (1996). Revision of the subgenus *Ceratitis* (*Pardalaspis*) Bezz, 1918 (Diptera, Tephritidae, Ceratitini). *Systematic Entomology* **21**: 15-26.

De Meyer M. (1998). Revision of the subgenus *Ceratitis* (*Ceratalaspis*) Hancock, 1918 (Diptera: Tephritidae, Ceratitini). *Bulletin of Entomological Research* **88**: 257-290.

De Meyer M. (2000). Systematic revision of the subgenus *Ceratitis* MacLeay *s. s* (Diptera: Tephritidae). *Zoological Journal of the Linnaean Society,* **128**: 439-467.

De Meyer M. (2001). On the identity of the Natal Fruit Fly, *Ceratitis rosa* (Diptera: Tephritidae). *Bulletin de l'Institut royal des Sciences Naturelles de Belgique, Entomologie* **71**: 55-62.

De Meyer M., Robertson M.P., Mansell M.W., Ekesi S., Tsuruta K., Mwaiko W., Vayssie`res J-F & Peterson A.T. (2009). Ecological niche and potential geographic distribution of the invasive fruit fly *Bactrocera invadens* (Diptera, Tephritidae) *Bulletin of Entomological Research*, Page 1 of 14.

Dieringer D. & Schlötterer C. (2003). Microsatellite Analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes*, **3**, 167-111.

Drew R.A.I. & Hancock D.L. (1994). The *Bactrocera dorsalis* complex of fruit flies (Diptera: Tephritidae: Dacinae) in asia. *Bulletin of Entomological Research*, Supplement Series **2**: 1-68.

Drew R. A. I. (1989). The taxonomy and distribution of tropical and subtropical Dacinae (Diptera: Tephritidae). Pp 9-14 in Robinson, A.S. & Hooper, G. (Eds) Fruit flies, their biology, natural enemies and control. Vol. **3A**. *Amsterdam, Elsevier.*

Drew R.A.I., Tsuruta K. & White. I.M. (2005). A new species of pest fruit fly (Diptera: Tephritidae: Dacinae) from Sri Lanka and Africa. *African Entomology*, **13**: 149-154.

Drew R.A.I., Raghu S. & Halcoop P. (2008). Bridging the morphological and biological species concepts: studies on the *Bactrocera dorsalis* (Hendel) complex (Diptera: Tephritidae: Dacinae) in South-east Asia. *Biological Journal of the Linnean Society* 93, 217–226.

Ebach M.C. & Holdrege C. (2005). DNA barcoding is no substitute for taxonomy. *Nature* 434, 697.

Eck R.V. & Dayhoff M.O. (1996). *Atlas of Protein Sequence and Structure*. National Biomedical Research Foundation, Silver Springs, Maryland.

Ekesi S.; Maniania N.K. & Lux S.A. (2002). Mortality in three African Tephritid fruit fly puparia and adults caused by the entomopathogenic fungi, *Metarhizium anisopliae* and *Beauveria bassiana*. *Biocontrol Science and Technology* **12**: 7-17.

Ekesi S., Nderitu P., Rwomushana I. (2006). Field infestation, life history and demographic parameters of the fruit fly *Bactrocera invadens* (Diptera: Tephritidae) in Africa. *Bulletin of Entomological Research,* **96**, 379-386.

Ekesi S. & Billah M. K. (2007). A field guide to the management of economically important Tephritid fruit flies in Africa. 2nd Edition, *icipe Science press.*

Erpenbeck D., Hooper J.N.A., Wörheide G. (2005). CO1 phylogenies in diploblasts and the 'Barcoding of Life' – are we sequencing a suboptimal partition? *Mol. Ecol. Notes* **1**, 1–4.

Evanno G., Regnaut S., Goudet J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L., Laval G., Schneider S. (2005). Arlequin version 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

Falush D., Stephens M., Pritchard J.K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567-1587.

Felsenstein J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**:783-791.

Felsenstein J. (2005). PHYLIP (Phylogeny Inference Package), Version 3.6. *Distributed by the Author*. Department of Genome Sciences, University of Washington, Seattle, Washington.

Folmer O., Black M., Hoeh W., Lutz R. & Vrijenhoek R. (1994). DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biological Biotechnology* **3**: 294–299.

French C. (2005). *The new invasive Bactrocera species*. pp. 19-20 *in Insect Pest Control Newsletter, No. 65*. International Atomic Energy Agency, Vienna, Austria.

Gasperi G., Bonizzoni M., Gomulski L. M. *et al.* (2002). Genetic differentiation, gene flow and the origin of infestations of the medfly, *Ceratitis capitata*. *Genetica*, **116**, 125–135.

Goudet J. (2001). FSTAT Version 2.9.3. a program to estimate and test gene diversities and fixation indices. http://www2. unil.ch/popgen/softwares/fstat.htm.

Greenway P.J. (1973). A classification of the vegetation of East Africa. Kirkia. **9**:1–68.

Griffiths C.J. (1993). The geological evolution of East Africa. 9–21. *in* Lovett, J. C. and S. K. Wasser, editors. eds. Biogeography and Ecology of the Rain Forests of Eastern Africa. Cambridge University Press. Cambridge. ix +. 341 pp.

Hajibabaei M., de Waard R., Ivanova V., Ratnasingham S., Dooh T., Kirk L., Mackie M. & Hebert P.D.N. (2005). Critical factors for assembling a high volume of DNA barcodes. Philosophical Transactions of the Royal Society (London) B Biological Science 360: 1959–1967.

Hajibabaei M. et al. (2006) DNA barcodes distinguish species of tropical Lepidoptera. *Proc. Natl. Acad. Sci.* U. S. A. 103, 968–971.

Hajibabaei M., Singer G.A., Hebert P.D.N. & Hickey D.A.C. (2007). DNA barcoding: how it complements taxonomy, molecular Phylogenetics and population genetics. *Science direct* 10:1-6.

Han H.Y. & McPheron B.A. (1997). Molecular phylogenetic study of Tephritidae (Insecta: Diptera) using partial sequences of the mitochondrial 16S ribosomal DNA. *Molecular Phylogenetic Evolution* 7: 17 – 32.

Hedrick P.W., Parker K.M. & Lee R.N. (2001). Using microsatellite and MHC variation to identify species, ESUs, and Mus in endangered Sonoran topminnow. *Molecular Ecology* **10**: 1399-1412.

Hebert P.D., Cywinska A., Ball S.L. & de Waard J.R. (2003a). Biological identifications through DNA barcodes. *Proceedings of the Royal Society* B 270: 313–321.

Hebert P.D.N., Ratnasingham S. & de Waard J.R. (2003b). Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. Lond. B Biol. Sci.* **270**(Suppl.): S96–S99.

KENYATTA UNIVERSITY LIBRARY

Hebert P.D., Penton E.H., Burns J.M., Janzen D.H. & Hallwachs W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. *Proceedings of the National Academy of Sciences* USA 101: 14 812–14 817.

Hebert P.D.N. & Gregory T.R. (2005). The promise of DNA barcoding for taxonomy. *Syst. Biol.* **54**, 852–859.

Heraty J.M. & Woolley J.B. (1993). Separate species or polymorphism:A recurring problem in Kapala (Hymenoptera: Eucharitidae). *Annals of the Entomological Society of America,* **86**(5): 517-531.

Hoffmeister T. (1992). Factors determining the structure and diversity of parasitoid complexes in tephritid fruit flies. *Oecologia* **89**: 288-297.

Holway D.A. & Suarez A.V. (1999). *Trends Ecol. Evol.* **14**, 328–330.

Hutter C.M., Schung M.D. & Aquadro C.F. (1998). Microsatellite variation in *Drosophila melanogaster* and *Drosophila simulans*: a reciprocal test of the ascertainment bias hypotesis. *Molecular Biology and Evolution,* **15**: 1620-1636.

Jarne P. & Lagoda P.J.L. (1996). Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution*, **11**, 424 – 429.

Jean-François Vayssières, Sam Korie & David Ayegnon (2009). Correlation of fruit fly (Diptera Tephritidae) infestation of major mango cultivars in Borgou (Benin) with abiotic and biotic factors and assessment of damage. *Crop protection*, **28**: 477-488.

Jeanmougin F., Thompson J.D., Gouy M., Higgins D.G. & Gibson T.J. (1998). Multiple Sequence Alignment with ClustalX. *Trends Biochem Sci.*, **23**: 403 – 405.

Jolicoeur P. & Mosimann J.E. (1960). Size and shape variation in the painted turtle.A principal component analysis. *Growth*, **24**: 339-354.

Kalinowski S.T., Taper M.L. & Marshall T.C. (2007). Revising how the computer program Cervus accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*, **16**, 1099–1006.

Kimani-Njogu S.W., Trostle M.K., Wharton R.A., Woolley J.B. & Raspi A. (2001). Biosystemics of the *Psyttalia concolor* species complex (Hymenoptera: Braconidae: Opiinae): The Identity of the populations attacking *Ceratitis capitata* (Diptera: Tephritidae) in coffee in Kenya. *Biological Control* **20**: 167-174.

Kimura M. (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*. **16**: 111–120.

Kinnear M.W., Bariana H.S., Sved J.A., Frommer M. (1998). Polymorphic microsatellite markers for population analysis of a tephritid pest species, *Bactrocera tryoni*. *Molecular Ecology*, **7**: 1489-1495.

Khamis F, Karam N, Guglielmino CR et al. (2008). Isolation and characterization of microsatellite markers in the newly discovered invasive fruit fly pest in Africa, *Bactrocera invadens* (Diptera: Tephritidae). *Molecular Ecology Resources*, **8**, 1509–1511.

Knapik E.W., Goodman A., Ekker M. *et al.* (1998). A microsatellite genetic linkage map for zebrafish (*Danio rerio*). *Nature Genetics*, **18**, 338 – 343.

Knipling E.F. (1992). Principles of insect parasitism analyzed from new perspectives. Practical implications for regulating insect populations by biological means. *United States Department of Agriculture, Agriculture Research Service, Agricultural Hand Book* **693**: 1-335.

Kshirsager A.M. (1972). Multivariate analysis. New York: Marcel Dekker, Inc.

Lambert D.M., Baker A., Huynen L., Haddrath O., Hebert P.D.N., Millar C.D. (2005). Is a large-scale DNA-based inventory of ancient life possible? *J. Hered.* **96**, 279–284.

Lawson A.E., McGuire D.J., Yeates D.K., Drew R.A.I., Clarke A.R. (2003). *Dorsalis*: an interactive identification tool to fruit flies of the *Bactrocera dorsalis* complex. CD-ROM Publication, Griffith University, Brisbane, Australia.

Lee C.E. (2002). Evolutionary genetics of invasive species. Trends Ecol Evol 17:386–391

Litt M. & Luty J.A. (1989). A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics*, **44**, 397 – 401.

Lodge D.M. (1993). *Trends Ecol. Evol.* **8,** 133–137.

Loxdale H.D. & Lushai G. (1998). Molecular markers in entomology. *Bulletin of Entomological Research*, **88:** 557 – 600.

Lux S.A., Copeland R.S. White I.M., Manrakhan A. & Billah M.K. (2003a). A New Invasive Fruit Fly Species from the *Bactrocera dorsalis* (Hendel) Group Detected in East Africa. *Insect Science and Its Application,* **23(4)**: 355–361.

Lux SA, Ekesi S, Dimbi S, Mohamed S & Billah MK (2003b) Mango-infesting fruit flies in Africa: perspectives and limitations of biological approaches to their management. Biological Control in IPM Systems in Africa (ed. by P Neuenschwander, C Borgemeister & J Langewald), pp. 277–293. CABI, Wallingford, UK.

Lux S.A., Ekesi S., Dimbi S., Mohamed S. & Billah M (2003). Mango-infesting fruit flies in Africa: Perspectives and limitations of biological approaches to their management. In *Biological Control in IPM Systems in Africa* (Edited by P. Neuenschwander, C. Borgemeister and J. Langewald). CABI, Wallingford.

Malacrida A.R., Marinoni F., Torti C., Gomulski L.M., Sebastiani F., Bonvivini C., Gasperi G., & Guglielmino C.R. (1998). Genetic aspects of the worldwide colonization process of *Ceratitis capitata*. *Journal of Heredity* **89**: 501-507.

Marcus L.F. (1990). *Traditional morphometrics*, pp. 77-122. In: Proceeding of the Michigan Morphometrics Workshop. Special Publication No. 2, University of Michigan. Museum of Zoology: Ann Arbor.

Mardulyn P. & Whitfield J.B. (1999). Phylogenetic signal in the COI, 16S, and 28S genes for inferring relationships among genera of Microgastrinae (Hymenoptera: Braconidae): evidence of a high diversification rate in this group of parasitoids. *Molecular Phylogenetics and Evolution.* 12: 282–294.

Marshall E. (2005). Taxonomy. Will DNA barcodes breath life into classification? *Science* (Washington, D.C.), 307: 1037.

McPheron B.A. & Steck G.J. (1996). Fruit Fly Pests: A world assesessment of their biology and management. St Lucie Press, Delray Beach, FL.

Meixner M.D., McPheron B.A., Silva J.G., Gasparich G.E., Sheppard W.S. (2002). The Mediterranean fruit fly in California: evidence for multiple introductions and persistent populations based on microsatellite and mitochondrial DNA variability. *Molecular Ecology*, **11**, 891–899.

Meyer C.P. & Paulay G. (2005). DNA barcoding: error rates based on comprehensive sampling. *PLoS Biol.* 3, e422.

Moritz C. & Brown W.M. (1987). Tandem duplications in animal mitochondrial DNAs : Variation in incidence and gene content among lizards . *Proc. Natl . Acad . Sci.* USA, **84**:7183-7187.

Moritz C. & Cicero C. (2004). DNA barcoding: promise and pitfalls. *PLoS Biol.* 2, 1529–1531.

Morrow J., Scott L., Congdon B., Yeates D., Frommer M. & Sved J. (2000). Close genetic similarity between two sympatric species of tephritid fruit fly reproductively isolated by mating time. *Evolution* **54**: 899-910.

Mwatawala M.W., White I.M., Maerere A.P., Senkondo F.J. & De Meyer M. (2004). A new invasive *Bactrocera* species (Diptera: Tephritidae) in Tanzania. *African Entomology*, 12: 154-156.

Mwatawala M.W., De Meyer M., Makundi R.H. & Maerere A.P. (2009). Host range and distribution of fruit-infesting pestiferous fruit flies (Diptera, Tephritidae) in selected areas of Central Tanzania. *Bulletin of Entomological Research*, Page 1 of 13.

Munro H.K. (1984). A taxonomic treatise on the Dacidae (Tephritidae: Diptera) of Africa. *Entomology Memoirs, Department of Agriculture & Water Supply, Republic of South Africa,* **61**: 1-313.

Nardi F., Carapelli A., Dallai R., Roderick G.K. & Frati F. (2005). Population structure and colonization history of the olive fly, *Bactrocera oleae* (Diptera, Tephritidae). *Molecular Ecology* **14**: 2729-2738.

Neff N.A. & Marcus L.F. (1980). Asurvey of multivariate methods for systematics. New York, 243 pp.

Nei M. & Kumar S. (2000). *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.

Nei M., Maruyama T., Chakraborty R. (1975). The bottleneck effect and genetic variability in populations. *Evolution*, **29**, 1–10.

Nei M., Tajima F., Tateno Y. (1983). Accuracy of estimated phylogenetic trees from molecular data. *Journal of Molecular Evolution*, **19**, 153-170.

Newell I.M. & Haramoto F.H. (1968). Biotic factors influencing populations of *acus dorsalis* in Hawaii. Proceedings of *Hawaiian Entomological Society* **20**: 81-139.

Norrbom A.L., Carroll L.E., Thompson F.C., White I.M. & Friedberg A. (1999). Systematic database of names. In: F.C. Thompson (Ed.) *Fruit Fly Expert Identification System Information Database*. Myia 9 (1998): 65-251.

Norrbom, A.L. (2002). The Diptera Site. Fruit fly (Diptera: Tephritidae). Available on: *http://www.sel.barc.usda.gov/diptera/tephriti.htm*

Omondi A.B., Obeng-Ofori D., Danquah E.Y. & Offei S.K. (2003). Genetic diversity and oviposition host preference of *Bemisia tabaci* associated with four crop hosts in Ghana. *Journal of the Ghana Science Association,* **6:** 105 – 116. *Parasite and Predators of Arthropod Pest and Weeds: A World Review* (C. P. Clausen, ed); U.S. Department of Agriculture Handbook No, **480**. *pp*: 320-335.

Paetkau D., Slade R., Burden M., Estoup A. (2004). Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology*, **13**, 55–65.

Page R.D.M. (1996). TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**: 357-358.

Peakall R. & Smouse P.E. (2005). Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. Australian National University, Canberra, Australia. http://www.anu.edu.au/BoZo/GenAlEx/.

Penton E.H., Hebert P.D.N. & Crease T.J. (2004). Mitochondrial DNA variation in North American populations of *Daphnia obtusa*: continentalism or cryptic endemism? *Mol. Ecol.* **13**, 97–107.

Piry S., Luikart G. & Cornuet J.M. (1999). Bottleneck: a computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity*, **90**, 502-503.

Piry S., Alapetite A., Cornuet J.M., Paetkau D., Baudouin L., Estoup A. (2004). Geneclass 2: a software for genetic assignment and first– generation migrant detection. *Journal of Heredity*, **95**, 536–539.

Pritchard J.K., Stephens M., Donnelly P. (2000). Inference of population structure using multilocus genotypic data. *Genetics*, **155**, 945-959.

Prokopy R.J. & Roitberg B.D. (1984). Foraging behaviour of true fruit flies. *American Scientist*, **72**: 41-49.

Purcell M.F. (1998). Contribution of biological control to integrated pest management of tephritid fruit flies in the tropics and subtropics. *Integrated Pest Management Reviews* **3**: 63-93.

Quinteiro J., Sotelo C.G., Rehbein H., Pryde S.E., Medina I., Perez-Martin R.I., Rey-Mendez M. & Mackiem I.M. (1998). Use of mtDNA direct polymerase chain reaction (PCR) sequencing and PCR-restriction fragment length polymorphism methodologies in species identification of canned tuna. *Journal of Agriculture and Food Chemistry*, **46**(4): 1662-1669.

Rannala B. & Mountain J.L. (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences, USA*, **94**, 9197–9221.

KENYATTA UNIVERSITY LIBRARY

Rao C.R. (1964). The use and interpretation of principal component analysis in applied research. *Sankhya* A, 39: 555 – 358.

Ratnasingham S. & Hebert P. (2007). BOLD: The Barcode of Life Data System. *Molecular Ecology Notes* **10**: 1-10.

Raymond M. & Rousset F. (1995). GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, 86, 248-249.

Rice W.R. (1989). Analysis tables of statistical tests. *Evolution*, 43, 223-225.

Rice R.E., Phillips P.A., Stewart-Leslie J. & Sibbett G.S. (2003). Olive fruit fly populations measured in central and southern California. California Agriculture, **57**, 122-127.

Richtsmeier J.T., Deleon V.B. & Lele S.R. (2002). The Promise of Geometric Morphometrics. *Yearbook of Physical Anthropology*, **45**: 63-91.

Roderick G. & Navajas M. (2003). Genotypes in novel environments: Genetics and evolution in biological control. *Nature Reviews Genetics* **4**: 889-899.

Roe A.D. & Sperling F.A.H. (2007). Patterns of evolution of mitochondrial cytochrome c oxidase I and II DNA and implications for DNA barcoding. *Molecular Phylogenetics and Evolution* 44, 325–345.

Rohlf F.J. & Bookstein F.L. (1990). Proceedings of the Michigan Morphometrics Workshop. *Special publication No. 2.* University of Michigan, Museum of Zoology: Ann Arbor.

Rozen S. & Skaletsky H.J. (2000). Primer 3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology*, pp. 365–386. Humana Press, Totowa, New Jersey.

Saitou N. & Nei M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

Saki A.K., Allendorf F.W., Holt J.S., Lodge D.M., Molofsky J. & With K.A. (2001). The population biology of invasive species. AnnuRev Ecol Syst 32:305–332.

SAS Institute (2000). *SAS user's guide*. Release 8.1, Cary, SAS Institute, USA.

Schug M.D., Hutter C.M., Wetterstrand K.A., Gaudette M.S., Mackay T.F.C. & Aquadro C.F. (1998a). The mutation rate of di-, tri- and tetranucleotide in *Drosophila melanogaster*. Molecular Biology and Evolution, 15: 1751-1760.

Schug M.D., Wetterstrand K.A., Gaudette M.S., Lim R.H., Hutter C.M. & Aquadro C.F. (1998b). The distribution and frequency of microsatellite loci in *Drosophila melanogaster*. Molecular Ecology, 7:57-70.

Seewooruthun S.I., Sookar P., Permalloo S., Joomaye A., Alleck M., Gungah B. & Soonnoo A.R. (1998). *An attempt of the eradication of the Oriental Fruit Fly,Bactrocera dorsalis (Hendel) from Mauritius*, pp. 211-218. *In*: Lalouette, J. A., Bachraz, D. Y., Sukurdeep, N. & Seebaluck, B. D. (Eds.). Proceedings of the Second Annual Meeting of Agricultural Scientists, 12-13 August, 1997, Réduit, Mauritius: *Food and Agriculture Research Council.*

Shearman D.C.A., Gilchrist A.S., Crisafulli D., Graham G., Lange C. & Frommer M. (2006). Microsatellite markers for the pest fruit fly, *Bactrocera Papayae* (Diptera: Tephritidae) and other *Bactrocera* species. *Molecular Ecology Notes* **6**, 4 –7.

Sheppard W.S., Steck G.S. & McPheron B.A. (1992). Geographical populations of the medfly may be differentiated by mitochondrial DNA variation. *Experienta*, 48: 1010-1013.

Schlötterer C. & Tautz D. (1992). Slippage synthesis of simple sequence DNA. *Nucleic Acids Research*, **20**, 211 – 215.

Schuler G.D., Boguski M.S., Stewart E.A. *et al.* (1996). A gene map of the human genome. *Science*, **274**, 540 – 546.

Sivinski J.M., Calkins C.O., Baranowski R., Harris D., Brambila J., Diaz J., Burns R.E., Holler T. & Dodson G. (1996). Suppression of a Caribbean fruit fly *Anastrepha suspensa* (loew) (Tephritidae: Diptera) population through augmented releases of the parasitoid *Diachasmimorpha longicaudata* (Ashmead) (Hymenoptera: Braconidae). *Biological Control* **6**: 177-185.

Smith M.A., Smith M.A., Woodley N.E., Janzen D.H., Hallwachs W., Hebert P.D.N., (2006). DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). *Proc. Natl. Acad. Sci.* U. S. A. 103, 3657– 3662.

Sneath P.H. & Sokal R.R. (1973). *Numerical Taxonomy: The Principles and Practice of Numerical Classification.* W.H. Freeman, San Francisco. 573 pp.

Sokal R.R. & Rohlf F.J. (1995). Biometry: The principles and Practice of Statistics in Biological Research. 3rd Edition. W. H. Freeman & Company, USA. 887 pp.

Song S.D., Jing M.A. & Hughes J. (2006). Polymorphic microsatellite DNA markers in *Bactrocera cacuminata* (Hering) (Diptera: Tephritidae). Molecular Ecology Notes, 6: 47-49.

Sonvinco A., Manso F. & Quesada-Allue L.A. (1996). Discrimination between the

immature stages of *Ceratitis capitata* and *Anastrepha fraterculus* (Diptera: Tephritidae) populations by random amplified polymorphic DNA polymerase chain reactiopn. *Journal of Economic Entomology* **89**: 1208-1212.

Sperling F.A.H. (2003). Butterfly molecular systematics. In: Boggs, C.L., Watt, W.B., Ehrlich, P.R. (Eds.), Butterflies: Ecology and Evolution Taking Flight. *University of Chicago Press, Chicago,* IL, pp. 431–458.

Tajima F. (1989). Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. *Genetics* **123**:585-595.

Tamura K., Dudley J., Nei M. & Kumar S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**:1596-1599.

Tan K.H. (Ed.) (2000). *Area-wide control of fruit flies and other insect pests.* Joint Proceeding of the International Conference in Area-Wide Control of Insect Pests, May 28-June 5 (1998), and the 5th International Symposiumon Fruit Flies of Economic Importance, June 1-5, 1998. Pernebit Universiti Sains Malaysia, Pulau Penang.782 pp.

Tautz D. (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research*, **17**, 6463 – 6471.

Thao M.L., Baumann L. & Baumann P. (2004). Organisation of the mitochondrial genomes of whiteflies, aphids and psyllids (Hemiptera, Stenorrhyncha) *BioMed Central Evolutionary Biology*, **4:** 25.

Thompson J.D., Gibson T.J., Plewniak F., Jeanmougin F. & Higgins D.G. (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research, 24:4876-4882.

Tsutsui N.D., Suarez A.V., Holway D.A., Case T.J. (2001). Relationship among native and introduced populations of the Argentine ant (*Linepithema humile*) and the source of introduced populations. Molecular Ecology, 10: 2151-2161.

Yu H., Frommer M., Robson M.K., Meats A.W., Shearman D.C., Sved J.A. (2001). Microsatellite analysis of the Queensland fruit fly Bactrocera tryoni (Diptera: Tephritidae) indicates spatial structuring: implications for population control. Bull Entomol Res 91:139–147.

Umeh V.C., Garcia L.E. & M. De Meyer (2008). Fruit flies of sweet oranges in Nigeria: species diversity, relative abundance and spread in major producing areas. *Fruits* 63: 145 – 153.

Umphrey G.J. (1996). Morphometric discrimination among sibling species in the *fulva-rudis-texana* complex of the ant genus *Aphaenogaster* (Hymenoptera: Formicidae). Canadan Journal of Zoology, **74**(3): 528-559.

Van Oosterhout C., Hutchinson W.F., Wills D.P., Shipley P. (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular ecology Notes*, 4, 535-538.

Weber J.L. & May P.E. (1989). Abundant class of human DNA polymorphism which can be typed using polymerase chain reaction. *American Journal of Human Genetics,* **44:** 388 – 396.

Weems H.V., Heppner J.B. & Nation J.L. (1999). Oriental fruit fly, *Bactrocera* (=Dacus) *dorsalis* (Hendel) (Insecta: Diptera: Tephritidae). Entomology and Nematology Department, Florida Cooperative Extension Services, Institute of Food and Agricultural Sciences,University of Florida, Circulars no. 21 and 303.

Weir B.S. & Cockerham C.C. (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358-1370.

Wessel G. (1997). Mango production and marketing in Kenya. 40 *pp.*

Wharton R.A. (1989). Classical Biological Control of fruit-infesting Tephritidae. In: *World Crop Pests. Fruit Flies: their biology, Natural Enemies and Control* (A.S. Robinson, and G. Hooper, eds); Elsevier, Amsterdam/New York. *pp*: 303-313.

Wharton R.A., Gilstrap F.E., Rhode R.H. & Fischel-M.M. (1981). Hymenopterous egg-pupal and larval-pupal parasitoids of *Ceratitis capitata* and *Anastrepha* spp. (Diptera: Tephritidae) in Costa Rica. *Entomophaga* **26**: 285-290.

White I.M. & Elson-Harris M.M. (1992). *Fruit flies of economic ignificance: their Identification and bionomics*. C.A.B International and Australian Centre for International Agricultural Research. 601 *pp.*

Will K.W. & Rubinoff D. (2004). Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics* **20,** 47–55.

Will K.W., Mishler B.D. & Wheeler Q.D. (2005). The perils of DNA barcoding and need for integrative taxonomy. *Syst. Biol.* **54**, 844–851.

Woolley J.B. & Browning H.W. (1987). Morphometric analysis of uniparental *Aphytis* reared from Chaff scale, *Pariatoria pergondii* Comstock, on Texas citrus (Hymenoptera: Aphelinidae, Homoptera: Diaspididae). *Proceedings of the Entomological Society of Washington*, **89**: 77-94.

Woolley J.B., Rose M. & Krauter P. (1994). Morphometric comparisons of *Aphytis* species in the *lingnanensis* group (Hymenoptera: Aphelinidae). Chapter 12, pp. 223-244,

In: Rosen, D., (Ed). Advances in *Aphytis* Research. Intercept Ltd., Andover, UK, 362 pages.

Zheng L., Benedict M.Q., Cornel A.J, Collins F.H., Kafatos F.C. (1996). An integrated genetic map of the African human malaria vector mosquito, *Anopheles gambiae*. Genetics, 143: 941-952.

APPENDICES

APPENDIX I

REAGENTS

1. EXTRACTION BUFFER
    - NaCl – 100 mM
    - Sucrose – 200 mM
    - EDTA – 50 mM
    - Tris-HCL pH 9.1 – 100 mM
    - SDS 0.5 %
    - Autoclaved and kept at room temperature.

2. GET Buffer
    - Glucose - 50 mM
    - EDTA – 10 mM
    - Tris-HCL pH 8 – 25 mM
    - Filter sterilize the reagent and store at 4 °C.

3. TBE Stock Buffer X5
    - Tris base – 54 g
    - Boric acid – 27.5 g
    - EDTA pH 8 0.5 M – 20 ml
    - Make to a final volume of 1000 ml.

4. TBE Gel preparation buffer
    - TBE X5 stock – 100 ml
    - Make to a final volume of 500 ml.

5. TBE Gel running buffer
    - TBE X5 stock – 200ml
    - Ethidium bromide – 40 µl.

- Make to a final volume of 1000 ml.

6. LB medium
   - Bacto® tryptone - 10 g
   - Bacto® yeast extract – 5 g
   - NaCL – 5 g
   - NaOH 1 M – 1 ml
   - Make to 1000 ml and autoclave the reagent.

7. LB Plate
   - Bacto® tryptone - 10 g
   - Bacto® yeast extract – 5 g
   - NaCL – 5 g
   - NaOH 1 M – 1 ml
   - Agar or Agarose – 15 g
   - Make to 1000 ml and autoclave the reagent.

8. SOC medium
   - Bacto® tryptone - 2.0 g
   - Bacto® yeast extract – 0.5 g
   - NaCL 1 M – 1 ml
   - KCL 1 M – 0.25 ml
   - Mg^{2+} stock, filter sterilized – 1 ml
   - Glucose 2 M, filtered sterilized – 1 ml
   - Make to 100 ml.
   - Stored at -20 °C.

9. X – Gal
   - 5-bromo-4-Chloro-3-indoyl-ß-D-galactoside – 100 mg
   - Dissolve in 2 ml of *N, N'*- dimethylformamide.

- Cover with aluminum foil and stored at -20 °C.

10. 20X SSC
    - NaCL 3 M (175 g/l)
    - Na_3citrate.$2H_2O$ 0.3 M (88 g/l)
    - Adjust to pH 7 with HCL 1 M.

12. 2X KGB buffer
    - Potassium glutamate (kept at 4 °C) – 200 mM
    - Tris-acetate pH 7.6 (adjust the pH with 0.5 M Trisma base) – 50 mM
    - Magnesium acetate – 20 mM
    - BSA – 100 µg/ml
    - Keep the above stock at 4°C.
    - Make 2-mercaptoethanol 10 mM stock and keep in a fume hood.
    - Add the 2-mercaptoethanol 10 mM just before use.

13. 1.5 % Agarose gel
    - 1.05 g Agarose
    - 70 ml TBE (gel preparation buffer)
    - 4 µl ethidium bromide.

14. 2 % Agarose gel
    - 1.4 g Agarose
    - 70 ml TBE (gel preparation buffer)
    - 4 µl ethidium bromide

APPENDIX II

SEQUENCES WITH MICROSATELLITE LOCI OF THE WORKING PRIMERS

1. Bi1

NNAGGGCGATTGGGCTCTAGATGATGCTCGAGCGGCGCAGTGTGATGGATAT
CTGCAGAATTCGCCCTTCTCTTGCTTACGCGTGGACTAACGCATATCCAATTA
CCACTGCTTACCGTTTGCTAAATCATGCACCCCTGGCGCCGTATTGCTTTTAA
TTAGTTTGACTCTTGACACTGGCCTCGTTTATTTAAGTTGATGCTCTTGGTCAC
ACACACACACACACATTCACATTCATGTAGTATAATGTCCTGCTCGTATTACA
GCTTTCTGACTTTGCATACCACTCGACTTTTTCACTTGCAGACGCGCCATACT
GGCCAACACTGATGTCTCCGGCCATACGCCTTGGCAGAGTATCTTCCAACAT
GATTACT

2. Bi2

TCNAACGTCNCNAGANACTCANNAACTTNTNAGAACANAAAGATNANCAAN
GCCTAATNCNNCNNANCCACNNNTNNANNANNGCANCATANNGNCAANTNT
GGCCCTCTAAATGCATGCTCGAGCGGCCGCCAGTGTGATGGATATCTGCAGA
ATTCGCCCTTCTCTTGCTTACGCGTGGACTAACGAGAGCACTCACTCAACCTT
AGACGCACACACACACACATGCAGGNAAATTGAAGTAAAAAAGTGTATTTTG
CATATTTCTCTTTTTTTTTGCTTTTAATAAGAATTAAAGTGTAAGAAGATTTTA
TTCGAAAAACTTCGCACAATTCAAGTGCTACTTGTTATGATTGCTGCTTTGCT
AACAACATTTGGAACCCAAGCCTCCCTCACTTTTTGCAAATAACTCTTTTATG
GCGCTTGNTTATCGTATTATTGTTGCTTTTTTGAACCANGTCAAAAAAATTGA
TGGAAAAAAATGTTTAAATGAAAAATTAAGAGTGGAAGTTCACTTTGCTNAA
ANTATGATTNGNAGTATTGAAGACCCGGTTTTTATCCAAA

3. Bi3

CTCTTGCTTACGCGTGGACTAACTTATCAAAGCAGCTGGGAAATCAAAGGCC
AAAACAGAAATGTTCACTTGTGGAACGCGCGAATTTCAAGCATTTTTTGTTTT
TAATCTGCCGCTGCGATTTGTTATGCAATCTCGTTGTTCTCCGTTGAAATACTG
TAAATGTGTGTGTGCTGCTCCGAGTTTTATCGCTTGTGCTTGGCCTTAAGACC
GGTTTAAAGGTTATTGCTTTTATTCAAACACAAGTCATAATTCAGACGCCAAT
TTTTTGTTCAAGAATTTATGCGTTCAACGGTAACAGATAAACTCCCGTTTAAG
CCGATTATTAAAGTGGNTTTGGTAGTAATAAATTCCCCTTTGTTNTGGAAAGA
AAATTAACGATTTGGCGCAAGCN

4. Bi4

ACATATACGCTGGAGTCACACACATACACCAACACAAGCGCACAAAGTAGG
GGAAATGCAGTTGAATTGCATGCGAGCGCTGCTCAATTCATGCATGTTTGTTT
GCATGTCTGTACATGTGTTCGTCGGCATGTTCATGCCTCAGCTGCGGCACTCG
CATGCTTGTAGTCATAAAAATTCATATCACATTCCATACGTTTGTCAACAGCA
TAACATACTCGCTCACCGCTGGCTCTCACATGCGCTCTCTCTCTCTGCTGGCTT
TTCGCACAAAACCGGCCGGCTATGTTGGCAACTACCCATGCACTATTCCATTA
AACTGGCAGCCAAAAATACCCGTAACTTGGGGGTGGGTAAACAATGGCATTG
NA

5. Bi5

ACAACAAAAATCTGCTGGCGTCTGCCGCTAATGTCGTTGCATTTGGGCTGCCA
CAATGCCACAAACGCTTTCGGCTGTCTCGCTCCGCGCTTGCTATCAGNCAGCC
AGTCAGTGTCTCGTCAATATTGTGTCATCAATCATTTATTCTCAACATAATGC
CATCACACACACAGACACATACATGTGTGCACGAAGTGGCATGGATTCACCG
AAAACAAACTCGCTTTGTAAACATGTCNAAGGAACAAAATATCGTTTAGCCT
AAAACGTCAAAATCACCGCTACTGTCAAGCGAAAATACACCAACAACAACA
ATAACTGTTTACTTTAGCTTACCATTACACTAAGCTTGTCGCATATATGAGAC
ACACACACACACACACACACATTTGGTGGCATTTAAGT

6. Bi6

GTCCACGCGACAAGTTCGACACAAAAGACAGCCAAGGCAAAGGCTTTGGCA
AGCGATAAAGCATCAAAGCAGCATACGCACACACCCACACAAGCACACGGC
ACAATCAGTATACTCATGCTATATATATGCTATTTGCTTACTTAGCTGATTTAT
TACGAGGAGTGCGCTTTGTCGTTGCATTAACACGAGTGTGATTTTCAAATATT
TCTAACCATAATTTGCTGCGACAACAAGCTACAGAAGGGTTAGTCCACGCGT
AAGCAAGAGAAGGGCGAATTCCAGCACACTGGCGGCCGTTACTAGTGGATCC
GAGCTCGGTANCAAGCTTGGCGTATCATGGTCATAGCTGTTCCTGTGTGGAA
ATGTTATCCGCTCACAATNC

7. Bi7

ACTCGCGCGTGTATGTATGCCTACTAATATCTACAGAGTTTCGTCTTCTCAAA
CTCGCTCTTCATTCAATCCACTTTTCTTCAAACTAAATTCACCAACACACACA
CAACACATGCTGCTTATGTCTCGTATGTGCATTGTTTGTGCAAATTTATTTTGC
CACTTAACGTGTCGCCCGCTTAATTCCGCAGCTTTAAAGTGCGTTAGTTTAAA
ATTTGCATAAAATTCGTGGCAAACGTGTTCTCAACACTGAATATATGAGGAT
GAAGAGCTGAAAGTAGACTACTTACTTTGTAGTCCACGCGTAAGCAAGAGAA
GGGCGAATTCCAGCACACTGGCGGCCGTTACTAGTGGATCCGAGCTCGGTAC
CAAGCTTGGCGTATCA

8. Bi8

ACTCAACAACAAGTAATGGTGACAAACAGCAAACCGACAAACAAGTGCAGC
AAAGACACGCGCATACACACACACATACACCTCAGTGCAAAGCGCATGAGT
GGAGGTGCCGGAAAGTGAGAAAAATTACGCTGCAACGTAGATCTGCGGTGA
AATGATAAATGAACGCCTCATGATGTGATAAAAATGGAATTAACATGACACC
GGAGCGGCAGCGCGCTTTGACAGCACCGCTCAACAGCGCAACAATGCTTCCC
CCTACCAGCAGCACCTTCTAATGCTCAGTCACCAGCAGGCGTACTTAGCCACT
CGGTCGGTGCAAGCGAGGTAGCTTGTTAATTATATCGGAATGCTGCGAACAC
TACGCTGCTGCTGTCATTGCCATTTCCTATGTATCCACGCGTAACAAGAGAAG
GCGAATTCACACCTGCGGCGTtCTAGTGGATCGACTC

9. Bi9

ACTCACGAAAATTAGGCCCGACTGCCGACAATTGACGTGCTTAATCAACCCT
TAAATAACGCTAATTTGTTGCGAACCCTCACCTCTTCAGCATCGCCAACGTCG
CCCGAGCAGCACACTCGTAACCGCAACGTAAACAAAAGCCAAATGCTAGAA
ATTATCACACAATGCCATGCCCGCCGGCGCTGCTCGTAAACATCTACATATTT

KENYATTA UNIVERSITY LIBRARY

ACATGTATGTGTATGTGTGTGTGTGCTGTTGCTTTGTTGTGTCGAAACTAGGT
GTGAATCCAAGTGTTTGCCCGCTTGATCGGTGCGTCCCTAGGGCCAAGAGCG
TGAATTTCAATTTATTATGCCTTATCGCCTCTAATGGCTGAAACGCTTGCGGC
GACCACGCAAAATTTCACAAGCTTACGTGCTTCCTGTGTGGGTTGCTGCATAA
ACTTACTATGCCAATG

10. Bi10

ACAACCAACATAAAAATAGATCAACCATGCCTACGGAAGAGAAATCTCTAAA
TTGCCCATTAAGCTGAGCACTCTTAAAACAGTAGCGAGTGAGCGCCAAAGAG
TCTGCTCAGCACATAGCACAAGTATGGACTCGGCTTTTATCCAAGATGCTCAA
ACGCGCCCACTTTTCCACCAAAACAGTGAGTGTAAAAATGTTGACAATCGAG
CAGATCACTGAGCAGTCGCTCAGCGCATGTGTGATTGTGTGTGTGTGTGTGCT
TGCTGGCTGATAAGTTGGAGAGTGGAACTGATCTGCACTTTATTTGGTTGTAC
AAAGCAGCTGATCGTGAAAGATTTGCTACTCGCGGCAATGCGGAAAGTAACA
AAACTATTTTTAGCTCTTGTCGTTGTTTTCGATCTGCAGCGAGCATAATTGCTC
TCGTTAGTCCCGCGTAACAAGAAAGGGG

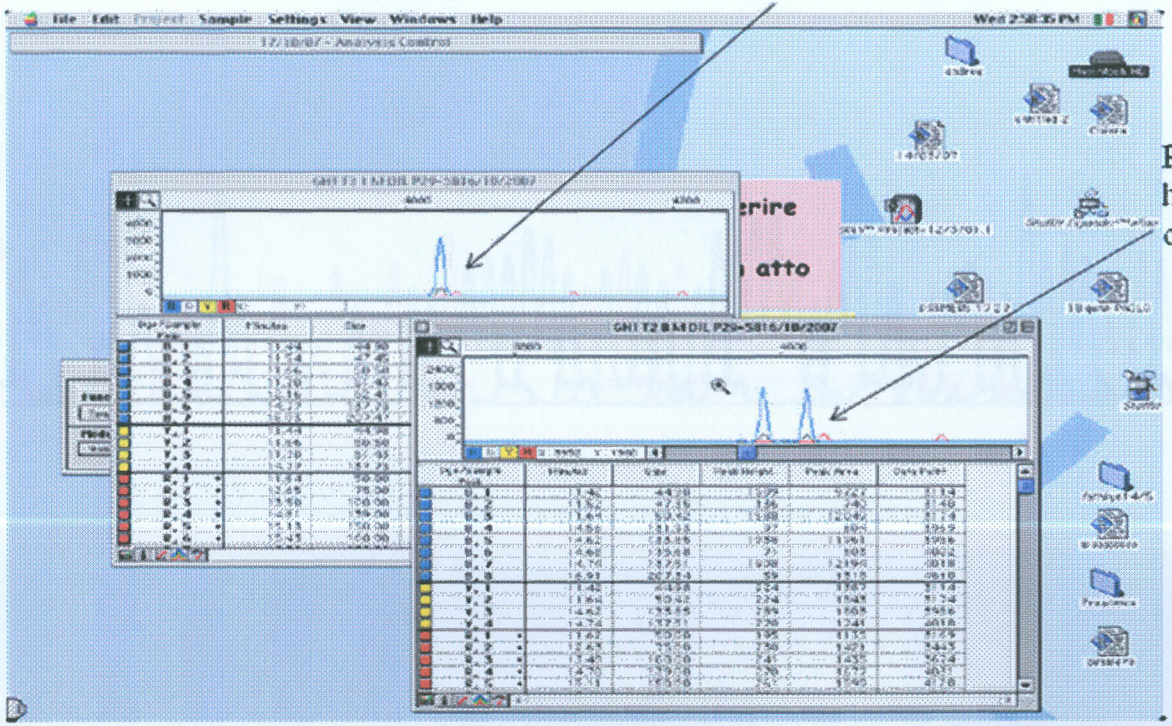
11. Bi11

ACACATCGAACAATACGGACTGAGGTCATTGTGAAATTTGTGTAGATAGCCT
CTGAAGCGCTCGCCCGCTTATATTTGGGTCATGGGGAAATCCAGGTCCCCATG
TTGTCTTCTTATCCACGTATGGATGTGAGGTATAAGTCTGTGGGTTCACCGTC
CTTTAATTGAGCTTTGCCACCGTTGCTGCCATATAGTTGTGCTCTTCATTCTCT
CTTCTCTCTTTTGGATTAATGTAATTGGCGCTGATAGCTCATAATAATCCCGT
GGGCTCGTTACCCTGGATGTCTATGGGCGGCATACAAT

APPENDIX III

Genescan output

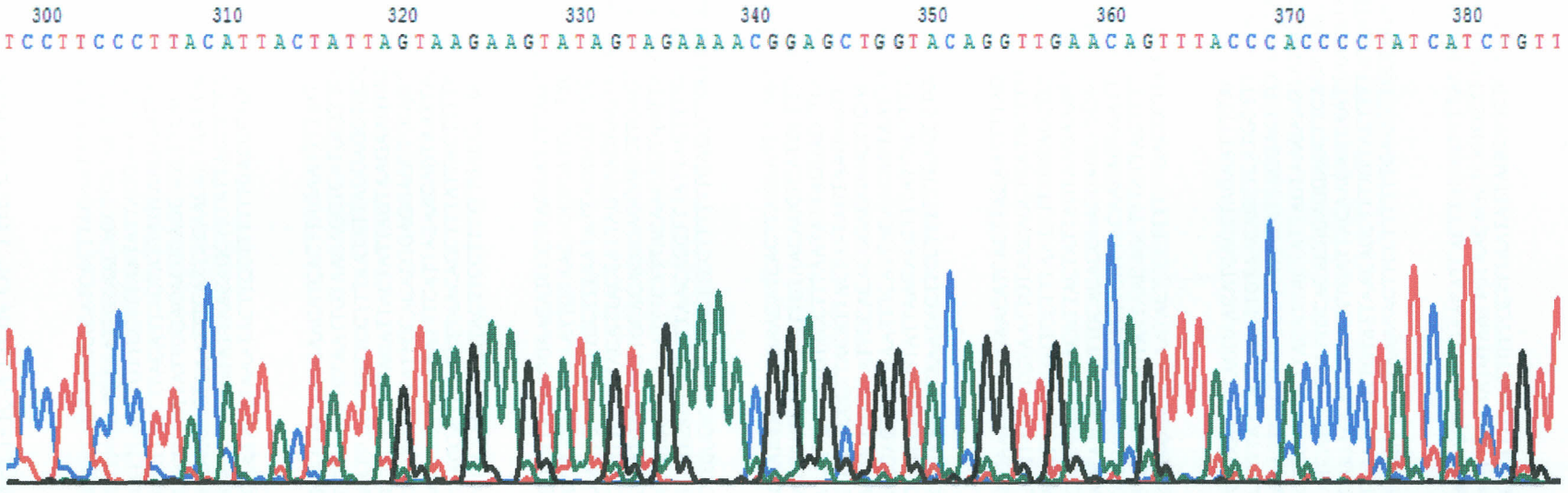
Profile showing a homozygote individual on a particular loci



Profile showing a heterozygote individual on a particular loci

APPENDIX IV

Chromatogram obtained from sequencing of the CO1 region for the Barcoding analysis



APPENDIX V

SEQUENCES OBTAINED BY AMPLIFICATION OF THE CO1 REGION

>BiNgT2.1
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCACTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGGGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiNgT2.2
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCACTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGGGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiNgT2.3
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGACTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTGTCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiNgT2.4
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCACTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGGGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiNgT2.5
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCACTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGGGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiNgT3.1
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCACTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGGGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiNgT3.2
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCACTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGGGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC

GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiNgT3.3
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGACTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTGTCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiNgT3.4
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGACTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTGTCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiNgT3.5
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGACTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTGTCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiNgT4.1
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCACTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGGGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiNgT4.2
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCACTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGGGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiNgT4.3
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCACTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGGGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiNgT4.4
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCACTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGGGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC

GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiNgT4.5
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCACTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGGGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiUgT2.1
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGACTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTGTCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiUgT2.2
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGACTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTGTCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiUgT2.3
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCACTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGGGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiUgT2.4
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCACTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGGGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiUgT2.5
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCACTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGGGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiUgT3.1
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC

GAATGAATAATATAAGATTTTGACTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTGTCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiUgT3.2
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGACTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTGTCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiUgT3.3
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGACTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTGTCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiUgT3.4
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGACTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTGTCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiUgT3.5
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGACTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTGTCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiUgT4.1
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCACTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGGGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTGTCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiUgT4.2
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGACTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTGTCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiUgT4.3
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC

GAATGAATAATATAAGATTTTGACTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTGTCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiUgT4.4
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGACTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTGTCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiUgT4.5
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGACTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTGTCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiZrT2.1
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATAAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCAGGTATCTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiZrT2.2
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCTCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGGGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiZrT2.3
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCTCTTAGAATTTTAGTTCGAGCTGAACTCGGTC
ATCCAGGAGCATTAATTGGAGACGACCAAATTTATAATGTAATTGTAACAGCCCATGCCTTCGTAATAATTTTCTTTATA
GTTATACCAATTATAATTGGTGGATTCGGAAATTGACTTGTTCCCCTAATATTAGGTGCTCCTGATATAGCTTTTCCACG
AATAAACAATATAAGATTTTGATTATTACCTCCTTCCCTTACACTACTATTAGTGAGAAGCATAGTAGAAAATGGGGCT
GGTACAGGTTGAACAGTTTACCCACCCCTATCATCGGTTATTGCACACGGAGGAGCTTCAGTTGATCTAGCTATTTTCTC
ACTCCACTTAGCCGGTATCTCTTCAATTTTAGGGGCAGTTAATTTCATTACAACAGTAATTAATATACGATCAACAGGA
ATCTCTTTTGATCGAATACCTCTTTTCGTATGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTTA
GCAGGAGCTATTACTATATTATTAACAGATCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGACCCTA
TTCTTTACCAACATTTATTT
>BiZrT2.4
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCTCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCCTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTCCACTTAGCGGGTATCTCCTCAATTTTAGGAGCAGTTAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCTCCTTTGATCGAATACCTCTTTACGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiZrT2.5
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCTCTTAGAATTTTAGTTCGAGCTGAACTCGGTC
ATCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCCTTCGTAATAATTTTCTTTATA
GTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCCTTAATATTAGGAGCTCCCGATATAGCATTTCCACG

AATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTGAGAAGTATAGTAGAAAACGGAGCT
GGTACAGGTTGAACAGTTTACCCACCCCTATCATCGGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTTC
ACTTCACTTAGCGGGTATCTCCTCAATTTTAGGAGCAGTTAATTTCATTACAACAGTAATTAATATACGATCGACAGGA
ATCTCCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTTA
GCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCTA
TTCTTTACCAACATTTATTT

>BiZrT3.1

AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCACTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGGGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT

>BiZrT3.2

AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCACTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGGGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT

>BiZrT3.3

AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTTGGAACATCTCTTAGAATTTTAGTTCGAGCTGAACTCGGTC
ATCCAGGAGCGTTAATTGGAGACGACCAAATTTATAATGTAATTGTAACAGCCCATGCCTTCGTAATAATTTTCTTTATA
GTTATACCAATTATAATTGGTGGATTCGGAAATTGACTTGTTCCCCTAATATTAGGTGCTCCTGATATAGCTTTTCCACG
AATAAACAATATAAGATTTTGATTATTACCTCCTTCCCTTACACTACTATTAGTGAGAAGTATAGTAGAAAATGGAGCT
GGTACAGGTTGAACAGTTTACCCACCCCTATCATCGGTTATTGCACACGGAGGAGCTTCAGTTGATCTAGCTATTTTCTC
ACTCCACTTAGCCGGTATCTCTTCAATTTTAGGAGCAGTTAATTTCATTACAACAGTAATTAATATACGATCAACAGGA
ATCTCTTTTGATCGAATACCTCTTTTCGTATGAGCAGTTGTATTAACAGCCTTATTACTTTTATTATCATTACCAGTTTTA
GCAGGAGCTATTACTATATTATTAACAGATCGAAACTTAAATACTTCCTTTTTTGACCCCGCCGGAGGGGGAGACCCTA
TTCTTTACCAACATTTATTT

>BiZrT3.4

AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCCCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT

>BiZrT3.5

AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCACTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGGGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT

>BiZrT4.1

AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTTGGAACATCTCTTAGAATTTTAGTTCGAGCTGAACTCGGTC
ATCCAGGAGCGTTAATTGGAGACGACCAAATTTATAATGTAATTGTAACAGCCCATGCCTTCGTAATAATTTTCTTTATA
GTTATACCAATTATAATTGGTGGATTCGGAAATTGACTTGTTCCCCTAATATTAGGTGCTCCTGATATAGCTTTTCCACG
AATAAACAATATAAGATTTTGATTATTACCTCCTTCCCTTACACTACTATTAGTGAGAAGTATAGTAGAAAATGGAGCT
GGTACAGGTTGAACAGTTTACCCACCCCTATCATCGGTTATTGCACACGGAGGAGCTTCAGTTGATCTAGCTATTTTCTC
ACTCCACTTAGCCGGTATCTCTTCAATTTTAGGAGCAGTTAATTTCATTACAACAGTAATTAATATACGATCAACAGGA
ATCTCTTTTGATCGAATACCTCTTTTCGTATGAGCAGTTGTATTAACAGCCTTATTACTTTTATTATCATTACCAGTTTTA
GCAGGAGCTATTACTATATTATTAACAGATCGAAACTTAAATACTTCCTTTTTTGACCCCGCCGGAGGGGGAGACCCTA
TTCTTTACCAACATTTATTT

>BiZrT4.2

AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCACTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGGGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC

KENYATTA UNIVERSITY LIBRARY

GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiZrT4.3
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiZrT4.4
AACTTTATATTTTATCTTCGGGGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiZrT4.5
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGCATTTCCTCAATTTTAGGAGCAGTTAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTTTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATACTACTAACAGACCGGAACTTAAACACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiSLT2.1
AACTTTATATTTTATCTTCGGGGCTTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTTCGAGCTGAACTCGGTC
ACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTATA
GTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCACG
AATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGCT
GGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTTC
ACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGGA
ATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTTA
GCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCTA
TTCTTTACCAACATTTATTT
>BiSLT2.2
AACTTTATATTTTATCTTCGGGGCTTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>BiSLT3.1
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTTGGAACATCTCTTAGAATTTTAGTTCGAGCTGAACTGGGTC
ATCCAGGAGCGTTAATTGGAGACGACCAAATTTATAATGTAATTGTAACAGCCCATGCCTTCGTAATAATTTTCTTTATG
GTTATACCAATTATAATTGGTGGGTTCGGAAATTGACTTGTTCCCCTAATATTAGGTGCTCCTGATATAGCTTTTCCACG
AATAAACAATATAAGATTTTGATTATTACCTCCTTCCCTTACACTACTATTAGTGAGAAGTATAGTAGAAAATGGAGCT
GGTACAGGTTGAACAGTTTACCCACCCCTATCATCGGTTATTGCACACGGAGGAGCTTCAGTTGATCTAGCTATTTTCTC
ACTCCACTTAGCCGGTATCTCTTCAATTTTAGGAGCAGTTAATTTCATTACAACAGTAATTAATATACGATCAACAGGA
ATCTCTTTTGATCGAATACCTCTTTTCGTATGAGCAGTTGTATTAACAGCCTTATTACTTTTATTATCATTACCAGTTTTA
GCAGGAGCTATTACTATATTATTAACAGATCGAAACTTAAATACTTCCTTTTTTGACCCCGCCGGAGGGGGAGACCCTA
TTCTTTACCAACATTTATTT
>BiSLT3.2
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC

GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGACCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAACACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT

>BiSLT3.3

AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTTGGAACATCTCTTAGAATTTTAGTTCGAGCTGAACTGGGTC
ATCCAGGAGCGTTAATTGGAGACGACCAAATTTATAATGTAATTGTAACAGCCCATGCCTTCGTAATAATTTTCTTTATG
GTTATACCAATTATAATTGGTGGGTTCGGAAATTGACTTGTTCCCCTAATATTAGGTGCTCCTGATATAGCTTTTCCACG
AATAAACAATATAAGATTTTGATTATTACCTCCTTCCCTTACACTACTATTAGTGAGAAGTATAGTAGAAAATGGAGCT
GGTACAGGTTGAACAGTTTACCCACCCCTATCATCGGTTATTGCACACGGAGGAGCTTCAGTTGATCTAGCTATTTTCTC
ACTCCACTTAGCCGGTATCTCTTCAATTTTAGGAGCAGTTAATTTCATTACAACAGTAATTAATATACGATCAACAGGA
ATCTCTTTTGATCGAATACCTCTTTTCGTATGAGCAGTTGTATTAACAGCCTTATTACTTTTATTATCATTACCAGTTTTA
GCAGGAGCTATTACTATATTATTAACAGATCGAAACTTAAATACTTCCTTTTTTGACCCCGCCGGAGGGGGAGACCCTA
TTCTTTACCAACATTTATTT

>BiSLT3.4

AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTTGGAACATCTCTTAGAATTTTAGTTCGAGCTGAACTGGGTC
ATCCAGGAGCATTAATTGGAGACGACCAAATTTATAATGTAATTGTAACAGCCCATGCCTTCGTAATAATTTTCTTTATA
GTTATACCAATTATAATTGGTGGATTCGGAAATTGACTTGTTCCCCTAATATTAGGTGCTCCTGATATAGCTTTTCCACG
AATAAACAATATAAGATTTTGATTATTACCTCCTTCCCTTACACTACTATTAGTGAGAAGCATAGTAGAAAATGGGGCT
GGTACAGGTTGAACAGTTTACCCACCCCTATCATCGGTTATTGCACACGGAGGAGCTTCAGTTGATCTAGCTATTTTCTC
ACTCCACTTAGCCGGTATCTCTTCAATTTTAGGGGCAGTTAATTTCATTACAACAGTAATTAATATACGATCAACAGGA
ATCTCTTTTGATCGAATACCTCTTTTCGTATGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTTA
GCAGGAGCTATTACTATATTATTAACAGATCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGACCCTA
TTCTTTACCAACATTTATTT

>BiSLT3.5

AACTTTATATTTTATCTTCGGGGCTTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT

>BiSLT4.1

AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTTGGAACATCTCTTAGAATTTTAGTTCGAGCTGAACTGGGTC
ATCCAGGAGCATTAATTGGAGACGACCAAATTTATAATGTAATTGTAACAGCCCATGCCTTCGTAATAATTTTCTTTATA
GTTATACCAATTATAATTGGTGGATTCGGAAATTGACTTGTTCCCCTAATATTAGGTGCTCCTGATATAGCTTTTCCACG
AATAAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACACTACTATTAGTGAGAAGCATAGTAGAAAATGGAGCT
GGTACAGGTTGAACAGTTTACCCACCCCTATCATCGGTTATTGCACACGGAGGAGCTTCAGTTGATCTAGCTATTTTCTC
ACTCCACTTAGCCGGTATCTCTTCAATTTTAGGGGCAGTTAATTTCATTACAACAGTAATTAATATACGATCAACAGGA
ATCTCTTTTGATCGAATACCTCTTTTCGTATGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTTA
GCAGGAGCTATTACTATATTATTAACAGATCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGACCCTA
TTCTTTACCAACATTTATTT

>BiSLT4.2

AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTTGGAACATCTCTTAGAATTTTAGTTCGAGCTGAACTGGGTC
ATCCAGGGGCATTAATTGGAGATGACCAAATTTATAATGTAATTGTAACAGCCCATGCCTTCGTAATAATTTTCTTTATA
GTTATACCAATTATAATTGGTGGATTCGGAAATTGACTTGTTCCCCTAATATTAGGTGCTCCTGATATAGCTTTTCCACG
AATAAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACACTACTATTAGTGAGAAGCATAGTAGAAAATGGAGCT
GGTACAGGTTGAACAGTTTACCCACCCCTATCATCGGTTATTGCACACGGAGGAGCTTCAGTTGATCTAGCTATTTTCTC
ACTCCACTTAGCCGGTATCTCTTCAATTTTAGGGGCAGTTAATTTCATTACAACAGTAATCAATATACGATCAACAGGA
ATCTCTTTTGATCGAATACCTCTTTTCGTATGAGCAGTTGTATTAACAGCTTTATTACTTTTATTGTCATTACCAGTTTTA
GCGGGAGCTATTACTATATTATTAACAGATCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGACCCTA
TTCTTTACCAACATTTATTT

>BiSLT4.3

AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTTGGAACATCTCTTAGAATTTTAGTTCGAGCTGAACTGGGTC
ATCCAGGAGCATTAATTGGAGACGACCAAATTTATAATGTAATTGTAACAGCCCATGCCTTCGTAATAATTTTCTTTATA
GTTATACCAATTATAATTGGTGGATTCGGAAATTGACTTGTTCCCCTAATATTAGGTGCTCCTGATATAGCTTTTCCACG
AATAAACAATATAAGATTTTGATTATTACCTCCTTCCCTTACACTACTATTAGTGAGAAGCATAGTAGAAAATGGAGCT
GGTACAGGTTGAACAGTTTACCCACCCCTATCATCGGTTATTGCACACGGAGGAGCTTCAGTTGATCTAGCTATTTTCTC
ACTCCACTTAGCCGGTATCTCTTCAATTTTAGGGGCAGTTAATTTCATTACAACAGTAATTAATATACGATCAACAGGA
ATCTCTTTTGATCGAATACCTCTTTTCGTGTGAGCAGTTGTATTAACAGCTTTATTGCTTTTATTATCATTACCAGTTTTA
GCAGGAGCTATTACTATATTATTAACAGATCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGGGGAGATCCTA
TTCTTTACCAACATTTATTT

>BiSLT4.4

AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTTGGAACATCTCTTAGAATTTTAGTTCGAGCTGAACTGGGTC
ATCCAGGAGCATTAATTGGAGACGACCAAATTTATAATGTAATTGTAACAGCCCATGCCTTCGTAATAATTTTCTTTATA
GTTATACCAATTATAATTGGTGGATTCGGAAATTGACTTGTTCCCCTAATATTAGGTGCTCCTGATATAGCTTTTCCACG

AATAAACAATATAAGATTTTGATTATTACCTCCTTCCCTTACACTACTATTAGTGAGAAGCATAGTAGAAAATGGAGCT
GGTACAGGTTGAACAGTTTACCCACCCCTATCATCGGTTATTGCACACGGAGGAGCTTCAGTTGATCTAGCTATTTTCTC
ACTCCACTTAGCCGGTATCTCTTCAATTTTAGGGGCAGTTAATTTCATTACAACAGTAATTAATATACGATCAACGGGA
ATCTCTTTTGATCGAATACCTCTTTTCGTATGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTTA
GCGGGAGCTATTACTATATTATTAACAGATCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGGGGAGACCCTA
TTCTTTACCAACATTTATTT
>BiSLT4.5
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTTGGAACATCTCTTAGAATTTTAGTTCGAGCTGAACTGGGTC
ATCCAGGAGCATTAATTGGAGACGACCAAATTTATAATGTAATTGTAACAGCCCATGCCTTCGTAATAATTTTCTTTATA
GTTATACCAATTATAATTGGTGGATTCGGAAATTGACTTGTTCCCCTAATATTAGGTGCTCCTGATATAGCTTTTCCACG
AATAAACAATATAAGATTTTGATTATTACCTCCTTCCCTTACACTACTATTAGTGAGAAGCATAGTAGAAAATGGAGCT
GGTACAGGTTGAACAGTTTACCCACCCCTATCATCGGTTATTGCACACGGAGGAGCTTCAGTTGATCTAGCTATTTTCTC
ACTCCACTTAGCCGGTATCTCTTCAATTTTAGGGGCAGTTAATTTCATTACAACAGTAATTAATATACGATCAACAGGA
ATCTCTTTTGATCGAATACCTCTTTTCGTGTGAGCAGTTGTATTAACAGCTTTATTGCTTTTATTATCATTACCAGTTTTA
GCAGGAGCTATTACTATATTATTAACAGATCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGGGGAGATCCTA
TTCTTTACCAACATTTATTT
>Bcor1
AACCTTATATTTTATTTTCGGGGCCTGAGCAGGAATAGTTGGGACATCCCTTAGAATTTTAGTTCGTGCCGAGCTAGGA
CACCCCGGAGCACTAATTGGAGATGACCAAATTTATAATGTAATTGTAACAGCACATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCTATCATAATTGGAGGATTTGGAAATTGACTTGTCCCCCTAATACTGGGAGCTCCTGATATAGCATTTCCAC
GAATAAATAATATAAGATTTTGATTATTACCCCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAATGGAGC
TGGTACAGGTTGAACAGTTTACCCTCCCCTATCATCTGTTATTGCTCACGGAGGAGCTTCGGTTGATCTAGCTATTTTCT
CACTCCACTTAGCTGGTATTTCCTCAGTTTTAGGAGCAGTAAATTTTATCACAACCGTTATTAACATGCGATCGACAGGA
ATTTCATTTGACCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTGCTTCTATTATCACTACCAGTTTTA
GCAGGAGCTATCACTATACTATTAACAGATCGAAACTTAAATACTTCCTTTTTTGACCCTGCAGGAGGAGGAGACCCTA
TTCTCTACCAACACTTATTT
>Bcor2
AACCTTATATTTTATTTTCGGGGCCTGAGCAGGAATAGTTGGGACATCCCTTAGAATTTTAGTTCGTGCCGAACTAGGA
CACCCCGGAGCACTAATTGGAGATGATCAAATTTATAATGTAATTGTAACAGCACATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCTATCATAATTGGAGGATTTGGAAATTGACTTGTCCCCCTAATACTGGGAGCTCCTGATATAGCATTTCCAC
GAATAAATAATATAAGATTTTGATTATTACCCCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAATGGAGC
TGGTACAGGTTGAACAGTTTACCCTCCCCTATCATCTGTTATTGCTCACGGAGGAGCTTCGGTTGATCTAGCTATTTTCT
CACTCCACTTAGCCGGTATTTCCTCAATTTTAGGAGCAGTAAATTTTATCACAACCGTTATTAACATACGATCGACAGG
AATTTCATTTGACCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTCTATTATCACTACCAGTTTT
AGCAGGAGCTATCACTATACTATTAACAGATCGAAACTTAAATACTTCCTTTTTTGACCCTGCAGGAGGAGGAGACCCT
ATTCTCTATCAACACTTATTT
>Bcor3
AACCTTATATTTTATTTTCGGGGCCTGAGCAGGAATAGTTGGGACATCCCTTAGAATTTTAGTTCGTGCCGAACTAGGA
CACCCCGGAGCACTAATTGGAGATGATCAAATTTATAATGTAATTGTAACAGCACATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCTATCATAATTGGAGGATTTGGAAATTGACTTGTCCCCCTAATACTGGGAGCTCCTGATATAGCATTTCCAC
GAATAAATAATATAAGATTTTGATTATTACCCCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAATGGAGC
TGGTACAGGTTGAACAGTTTACCCTCCCCTATCATCTGTTATTGCTCACGGAGGAGCTTCGGTTGATCTAGCTATTTTCT
CACTCCACTTAGCTGGTATTTCCTCAATTTTAGGAGCAGTAAATTTTATCACAACCGTTATTAACATGCGATCGACAGGA
ATTTCATTTGACCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTGCTTCTATTATCACTACCAGTTTTA
GCAGGAGCTATCACTATACTATTAACAGATCGAAACTTAAATACTTCCTTTTTTGACCCTGCAGGAGGAGGAGACCCTA
TTCTCTACCAACACTTATTT
>Bcor4
AACCTTATATTTTATTTTCGGGGCCTGAGCAGGAATAGTTGGGACATCCCTTAGAATTTTAGTTCGTGCCGAACTAGGA
CACCCCGGAGCACTAATTGGAGATGATCAAATTTATAATGTAATTGTAACAGCACATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCTATCATAATTGGAGGGTTTGGAAATTGACTTGTCCCCCTAATACTGGGAGCTCCTGATATAGCATTTCCAC
GAATAAATAATATAAGATTTTGATTATTACCCCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAATGGAGC
TGGTACAGGTTGAACAGTTTACCCTCCCCTATCATCTGTTATTGCTCACGGAGGAGCTTCAGTTGATCTAGCTATTTTCT
CACTCCACTTAGCTGGTATTTCCTCAATTTTAGGAGCAGTAAATTTTATCACAACCGTTATTAACATGCGATCGACAGGA
ATTTCATTTGACCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTGCTTCTATTATCACTACCAGTTTTA
GCAGGAGCTATCACTATACTATTAACAGATCGAAACTTAAATACTTCCTTTTTTGACCCTGCGGGAGGAGGAGACCCTA
TTCTCTACCAACACTTATTT
>Bcor5
AACCTTATATTTTATTTTCGGGGCCTGAGCAGGAATAGTTGGGACATCCCTTAGAATTTTAGTTCGTGCCGAACTAGGA
CACCCCGGAGCACTAATTGGAGATGATCAAATTTATAATGTAATTGTAACAGCACATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCTATCATAATTGGAGGATTTGGAAATTGACTTGTCCCCCTAATACTGGGAGCTCCTGATATAGCATTTCCAC
GAATAAATAATATAAGATTTTGATTATTACCCCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAATGGAGC
TGGTACAGGTTGAACAGTTTACCCTCCCCTATCATCTGTTATTGCTCACGGAGGAGCTTCGGTTGATCTAGCTATTTTCT
CACTCCACTTAGCTGGTATTTCCTCAATTTTAGGAGCAGTAAATTTTATCACAACCGTTATTAACATGCGATCGACAGGA
ATTTCATTTGACCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTGCTTCTATTATCACTACCAGTTTTA
GCAGGAGCTATCACTATACTATTAACAGATCGAAACTTAAATACTTCCTTTTTTGACCCTGCAGGAGGAGGAGACCCTA
TTCTCTACCAACACTTATTT
>Bcor6
AACCTTATATTTTATTTTCGGGGCCTGAGCAGGAATAGTTGGGACATCCCTTAGAATTTTAGTTCGTGCCGAACTAGGA
CACCCCGGAGCACTAATTGGAGATGATCAAATTTATAATGTAATTGTAACAGCACATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCTATCATAATTGGAGGATTTGGAAATTGACTTGTCCCCCTAATACTGGGAGCTCCTGATATAGCATTTCCGC

GAATAAATAATATAAGATTTTGATTATTACCCCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAATGGAGC
TGGTACAGGTTGAACAGTTTACCCTCCCCTATCATCTGTTATTGCTCACGGAGGAGCTTCGGTTGATCTAGCTATTTTCT
CACTCCACTTAGCCGGTATTTCCTCAATTTTAGGAGCAGTAAATTTTATCACAACCGTTATTAACATACGATCGACAGG
AATTTCATTTGACCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTCTATTATCACTACCAGTTTT
AGCAGGAGCTATCACTATACTATTAACAGATCGAAACTTAAATACTTCCTTTTTTGACCCTGCAGGAGGAGGAGACCCT
ATTCTCTATCAACACTTATTT
>Bcor7
AACCTTATATTTTATTTTCGGGGCCTGAGCAGGAATAGTTGGGACATCCCTTAGAATTTTAGTTCGTGCCGAACTAGGA
CACCCCGGAGCACTAATTGGAGATGATCAAATTTATAATGTAATTGTAACAGCACATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCTATCATAATTGGAGGATTTGGAAATTGACTTGTCCCCCTAATACTGGGAGCTCCTGATATAGCATTTCCAC
GAATAAATAATATAAGATTTTGATTATTACCCCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAATGGAGC
TGGTACAGGTTGAACAGTTTACCCTCCCCTATCATCTGTTATTGCTCACGGAGGTGCTTCGGTTGATCTAGCTATTTTCT
CACTCCACTTAGCTGGTATTTCCTCAATTTTAGGAGCAGTAAATTTTATCACAACCGTTATTAACATACGATCGACAGGA
ATTTCATTTGACCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTGCTTCTATTATCACTACCAGTTTTA
GCAGGAGCTATCACTATACTATTAACAGATCGAAACTTAAATACTTCCTTTTTTGACCCTGCAGGAGGAGGAGACCCTA
TTCTCTACCAACACTTATTT
>Bcor8
AACCTTATATTTTATTTTCGGGGCCTGAGCAGGAATAGTTGGGACATCCCTTAGAATTTTAGTTCGTGCCGAGCTAGGA
CACCCCGGAGCACTAATTGGAGATGACCAAATTTATAATGTAATTGTAACAGCACATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCTATCATAATTGGAGGATTTGGAAATTGACTTGTCCCCCTAATACTGGGAGCTCCTGATATAGCATTTCCAC
GAATAAATAATATAAGATTTTGATTATTACCCCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAATGGAGC
TGGTACAGGTTGAACAGTTTACCCTCCCCTATCATCTGTTATTGCTCACGGAGGAGCTTCGGTTGATCTAGCTATTTTCT
CACTCCACTTAGCTGGTATTTCCTCAATTTTAGGAGCAGTAAATTTTATCACAACCGTTATTAACATACGATCGACAGGA
ATTTCATTTGACCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTGCTTCTATTATCACTACCAGTTTTA
GCAGGAGCTATCACTATACTATTAACAGATCGAAACTTAAATACTTCCTTTTTTGACCCTGCAGGAGGAGGAGACCCTA
TTCTCTACCAACACTTATTT
>Bcor9
AACCTTATATTTTATTTTCGGGGCCTGAGCAGGAATAGTTGGGACATCCCTTAGAATTTTAGTTCGTGCCGAGCTAGGA
CACCCCGGAGCACTAATTGGAGATGACCAAATTTATAATGTAATTGTAACAGCACATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCTATCATAATTGGAGGATTTGGAAATTGACTTGTCCCCCTAATACTGGGAGCTCCTGATATAGCATTTCCAC
GAATAAATAATATAAGATTTTGATTATTACCCCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAATGGAGC
TGGTACAGGTTGAACAGTTTACCCTCCCCTATCATCTGTTATTGCTCACGGAGGAGCTTCGGTTGATCTAGCTATTTTCT
CACTCCACTTAGCTGGTATTTCCTCAATTTTAGGAGCAGTAAATTTTATCACAACCGTTATTAACATACGATCGACAGGA
ATTTCATTTGACCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTGCTTCTATTATCACTACCAGTTTTA
GCAGGAGCTATCACTATACTATTAACAGATCGAAACTTAAATACTTCCTTTTTTGACCCTGCAGGAGGAGGAGACCCTA
TTCTCTACCAACACTTATTT
>Bcor10
AACCTTATATTTTATTTTCGGGGCCTGAGCAGGAATAGTCGGGACATCCCTTAGAATTTTAGTTCGTGCCGAACTAGGA
CACCCCGGAGCACTAATTGGAGATGATCAAATTTATAATGTAATTGTAACAGCACATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCTATCATAATTGGAGGATTTGGAAATTGACTTGTCCCCCTAATACTGGGAGCTCCTGATATAGCATTTCCAC
GAATAAATAATATAAGATTTTGATTATTACCCCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAATGGAGC
TGGTACAGGTTGAACAGTTTACCCTCCCCTATCATCTGTTATTGCTCACGGAGGAGCTTCGGTTGATCTAGCTATTTTCT
CACTCCACTTAGCCGGTATTTCCTCAATTTTAGGAGCAGTAAATTTTATCACAACCGTTATTAACATACGATCGACAGG
AATTTCATTTGACCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTCTATTATCACTACCAGTTTT
AGCAGGAGCTATCACTATACTATTAACAGATCGAAACTTAAATACTTCCTTTTTTGACCCTGCAGGAGGAGGAGACCCT
ATTCTCTATCAACACTTATTT
>Bcu1
AACATTATATTTTATTTTCGGAGCTTGAGCAGGTATAGTGGGAACATCTCTTAGAATCTTAGTCCGGGCAGAACTGGGT
CACCCAGGAGCTTTAATCGGAGATGATCAAATCTATAATGTCATCGTAACAGCTCATGCATTTGTTATGATTTTTTTCAT
AGTGATACCTATTATAATTGGAGGATTTGGAAATTGACTAGTACCCCTAATACTAGGAGCGCCAGATATAGCATTCCCT
CGAATGAATAATATAAGATTTTGATTATTACCTCCCTCTCTTACATTACTTTTAGTGAGCAGTATAGTAGAAAACGGAG
CTGGTACAGGTTGAACTGTTTATCCTCCCCTTTCATCAATTATCGCTCATGGTGGAGCCTCAGTTGATTTAGCTATTTTTT
CTCTACATTTAGCTGGTATTTCATCAATTTTAGGGGCCGTAAATTTCATTACTACAGTAATTAATATGCGATCAACAGGA
ATCACATTTGACCGGATACCTTTATTCGTTTGAGCTGTAGTATTGACAGCTCTTCTTTTACTTCTATCTCTACCTGTATTA
GCTGGAGCTATTACTATACTTTTAACAGACCGAAATTTAAACACCTCTTTCTTCGACCCGGCTGGTGGTGGAGACCCTAT
TTTATACCAACATTTATTT
>Bcu2
AACATTATATTTTATTTTCGGAGCTTGAGCAGGTATAGTGGGAACATCTCTTAGAATCTTAGTCCGGGCAGAACTGGGT
CACCCAGGAGCTTTAATCGGAGATGATCAAATCTATAATGTCATCGTAACAGCTCATGCATTTGTTATGATTTTTTTCAT
AGTGATACCTATTATAATTGGAGGATTTGGAAATTGACTAGTACCCCTAATACTAGGAGCGCCAGATATAGCATTCCCT
CGAATGAATAATATAAGATTTTGATTATTACCTCCCTCTCTTACATTACTTTTAGTGAGCAGTATAGTAGAAAACGGAG
CTGGTACAGGTTGAACTGTTTATCCTCCCCTTTCATCAATTATCGCTCATGGTGGAGCCTCAGTTGATTTAGCTATTTTTT
CTCTACATTTAGCTGGTATTTCATCAATTTTAGGGGCCGTAAATTTCATTACTACAGTAATTAATATGCGATCAACAGGA
ATCACATTTGACCGGATACCTTTATTCGTTTGAGCTGTAGTATTGACAGCTCTTCTTTTACTTCTATCTCTACCTGTATTA
GCTGGAGCTATTACTATACTTTTAACAGACCGAAATTTAAACACCTCTTTCTTCGACCCGGCTGGTGGTGGAGACCCTAT
TTTATACCAACATTTATTT
>Bcu3
AACATTATATTTTATTTTCGGAGCTTGAGCAGGTATAGTGGGAACATCTCTTAGAATCTTAGTCCGGGCAGAACTGGGT
CACCCAGGAGCTTTAATCGGAGATGATCAAATCTATAATGTCATCGTAACAGCTCATGCATTTGTTATGATTTTTTTCAT
AGTGATACCTATTATAATTGGAGGATTTGGAAATTGACTAGTACCCCTAATACTAGGAGCGCCAGATATAGCATTCCCT

KENYATTA UNIVERSITY

CGAATGAATAATATAAGATTTTGATTATTACCTCCCTCTCTTACATTACTTTTAGTGAGCAGTATAGTAGAAAACGGAG
CTGGTACAGGTTGAACTGTTTATCCTCCCCTTTCATCAATTATCGCTCATGGTGGAGCCTCAGTTGATTTAGCTATTTTTT
CTCTACATTTAGCTGGTATTTCATCAATTTTAGGGGCCGTAAATTTCATTACTACAGTAATTAATATGCGATCAACAGGA
ATCACATTTGACCGGATACCTTTATTCGTTTGAGCTGTAGTATTGACAGCTCTTCTTTTACTTCTATCTCTACCTGTATTA
GCTGGAGCTATTACTATACTTTTAACAGACCGAAATTTAAACACCTCTTTCTTCGACCCGGCTGGTGGTGGAGACCCTAT
TTTATACCAACATTTATTT

>Bcu4

AACATTATATTTTATTTTCGGAGCTTGAGCAGGTATAGTGGGAACATCTCTTAGAATCTTAGTCCGGGCAGAACTGGGT
CACCCAGGAGCTTTAATCGGAGATGATCAAATCTATAATGTCATCGTAACAGCTCATGCATTTGTTATGATTTTTTTCAT
AGTGATACCTATTATAATTGGAGGATTTGGAAATTGACTAGTACCCCTAATACTAGGAGCGCCAGATATAGCATTCCCT
CGAATGAATAATATAAGATTTTGATTATTACCTCCCTCTCTTACATTACTTTTAGTGAGCAGTATAGTAGAAAACGGAG
CTGGTACAGGTTGAACTGTTTATCCTCCCCTTTCATCAATTATCGCTCATGGTGGAGCCTCAGTTGATTTAGCTATTTTTT
CTCTACATTTAGCTGGTATTTCATCAATTTTAGGGGCCGTAAATTTCATTACTACAGTAATTAATATGCGATCAACAGGA
ATCACATTTGACCGGATACCTTTATTCGTTTGAGCTGTAGTATTGACAGCTCTTCTTTTACTTCTATCTCTACCTGTATTA
GCTGGAGCTATTACTATACTTTTAACAGACCGAAATTTAAACACCTCTTTCTTCGACCCGGCTGGTGGTGGAGACCCTAT
TTTATACCAACATTTATTT

>Bcu5

AACATTATATTTTATTTTCGGAGCTTGAGCAGGTATAGTGGGAACATCTCTTAGAATCTTAGTCCGGGCAGAACTGGGT
CACCCAGGAGCTTTAATCGGAGATGATCAAATCTATAATGTCATCGTAACAGCTCATGCATTTGTTATGATTTTTTTCAT
AGTGATACCTATTATAATTGGAGGATTTGGAAATTGACTAGTACCCCTAATACTAGGAGCGCCAGATATAGCATTCCCT
CGAATGAATAATATAAGATTTTGATTATTACCTCCCTCTCTTACATTACTTTTAGTGAGCAGTATAGTAGAAAACGGAG
CTGGTACAGGTTGAACTGTTTATCCTCCCCTTTCATCAATTATCGCTCATGGTGGAGCCTCAGTTGATTTAGCTATTTTTT
CTCTACATTTAGCTGGTATTTCATCAATTTTAGGGGCCGTAAATTTCATTACTACAGTAATTAATATGCGATCAACAGGA
ATCACATTTGACCGGATACCTTTATTCGTTTGAGCTGTAGTATTGACAGCTCTTCTTTTACTTCTATCTCTACCTGTATTA
GCTGGAGCTATTACTATACTTTTAACAGACCGAAATTTAAACACCTCTTTCTTCGACCCGGCTGGTGGTGGAGACCCTAT
TTTATACCAACATTTATTT

>Bcu6

AACATTATATTTTATTTTCGGAGCTTGAGCAGGTATAGTGGGAACATCTCTTAGAATCTTAGTCCGGGCAGAACTGGGT
CACCCAGGAGCTTTAATCGGAGATGATCAAATCTATAATGTCATCGTAACAGCTCATGCATTTGTTATGATTTTTTTCAT
AGTGATACCTATTATAATTGGAGGATTTGGAAATTGACTAGTACCCCTAATACTAGGAGCGCCAGATATAGCATTCCCT
CGAATGAATAATATAAGATTTTGATTATTACCTCCCTCTCTTACATTACTTTTAGTGAGCAGTATAGTAGAAAACGGAG
CTGGTACAGGTTGAACTGTTTATCCTCCCCTTTCATCAATTATCGCTCATGGTGGAGCCTCAGTTGATTTAGCTATTTTTT
CTCTACATTTAGCTGGTATTTCATCAATTTTAGGGGCCGTAAATTTCATTACTACAGTAATTAATATGCGATCAACAGGA
ATCACATTTGACCGGATACCTTTATTCGTTTGAGCTGTAGTATTGACAGCTCTTCTTTTACTTCTATCTCTACCTGTATTA
GCTGGAGCTATTACTATACTTTTAACAGACCGAAATTTAAACACCTCTTTCTTCGACCCGGCTGGTGGTGGAGACCCTAT
TTTATACCAACATTTATTT

>Bcu7

AACATTATATTTTATTTTCGGAGCTTGAGCAGGTATAGTGGGAACATCTCTTAGAATCTTAGTCCGGGCAGAACTGGGT
CACCCAGGAGCTTTAATCGGAGATGATCAAATCTATAATGTCATCGTAACAGCTCATGCATTTGTTATGATTTTTTTCAT
AGTGATACCTATTATAATTGGAGGATTTGGAAATTGACTAGTACCCCTAATACTAGGAGCGCCAGATATAGCATTCCCT
CGAATGAATAATATAAGATTTTGATTATTACCTCCCTCTCTTACATTACTTTTAGTGAGCAGTATAGTAGAAAACGGAG
CTGGTACAGGTTGAACTGTTTATCCTCCCCTTTCATCAATTATCGCTCATGGTGGAGCCTCAGTTGATTTAGCTATTTTTT
CTCTACATTTAGCTGGTATTTCATCAATTTTAGGGGCCGTAAATTTCATTACTACAGTAATTAATATGCGATCAACAGGA
ATCACATTTGACCGGATACCTTTATTCGTTTGAGCTGTAGTATTGACAGCTCTTCTTTTACTTCTATCTCTACCTGTATTA
GCTGGAGCTATTACTATACTTTTAACAGACCGAAATTTAAACACCTCTTTCTTCGACCCGGCTGGTGGTGGAGACCCTAT
TTTATACCAACATTTATTT

>Bcu8

AACATTATATTTTATTTTCGGAGCTTGAGCAGGTATAGTGGGAACATCTCTTAGAATCTTAGTCCGGGCAGAACTGGGT
CACCCAGGAGCTTTAATCGGAGATGATCAAATCTATAATGTCATCGTAACAGCTCATGCATTTGTTATGATTTTTTTCAT
AGTGATACCTATTATAATTGGAGGATTTGGAAATTGACTAGTACCCCTAATACTAGGAGCGCCAGATATAGCATTCCCT
CGAATGAATAATATAAGATTTTGATTATTACCTCCCTCTCTTACATTACTTTTAGTGAGCAGTATAGTAGAAAACGGAG
CTGGTACAGGTTGAACTGTTTATCCTCCCCTTTCATCAATTATCGCTCATGGTGGAGCCTCAGTTGATTTAGCTATTTTTT
CTCTACATTTAGCTGGTATTTCATCAATTTTAGGGGCCGTAAATTTCATTACTACAGTAATTAATATGCGATCAACAGGA
ATCACATTTGACCGGATACCTTTATTCGTTTGAGCTGTAGTATTGACAGCTCTTCTTTTACTTCTATCTCTACCTGTATTA
GCTGGAGCTATTACTATACTTTTAACAGACCGAAATTTAAACACCTCTTTCTTCGACCCGGCTGGTGGTGGAGACCCTAT
TTTATACCAACATTTATTT

>Bcu9

AACATTATATTTTATTTTCGGAGCTTGAGCAGGTATAGTGGGAACATCTCTTAGAATCTTAGTCCGGGCAGAACTGGGT
CACCCAGGAGCTTTAATCGGAGATGATCAAATCTATAATGTCATCGTAACAGCTCATGCATTTGTTATGATTTTTTTCAT
AGTGATACCTATTATAATTGGAGGATTTGGAAATTGACTAGTACCCCTAATACTAGGAGCGCCAGATATAGCATTCCCT
CGAATGAATAATATAAGATTTTGATTATTACCTCCCTCTCTTACATTACTTTTAGTGAGCAGTATAGTAGAAAACGGAG
CTGGTACAGGTTGAACTGTTTATCCTCCCCTTTCATCAATTATCGCTCATGGTGGAGCCTCAGTTGATTTAGCTATTTTTT
CTCTACATTTAGCTGGTATTTCATCAATTTTAGGGGCCGTAAATTTCATTACTACAGTAATTAATATGCGATCAACAGGA
ATCACATTTGACCGGATACCTTTATTCGTTTGAGCTGTAGTATTGACAGCTCTTCTTTTACTTCTATCTCTACCTGTATTA
GCTGGAGCTATTACTATACTTTTAACAGACCGAAATTTAAACACCTCTTTCTTCGACCCGGCTGGTGGTGGAGACCCTAT
TTTATACCAACATTTATTT

>Bcu10

AACATTATATTTTATTTTCGGAGCTTGAGCAGGTATAGTGGGAACATCTCTTAGAATCTTAGTCCGGGCAGAACTGGGT
CACCCAGGAGCTTTAATCGGAGATGATCAAATCTATAATGTCATCGTAACAGCTCATGCATTTGTTATGATTTTTTTCAT
AGTGATACCTATTATAATTGGAGGATTTGGAAATTGACTAGTACCCCTAATACTAGGAGCGCCAGATATAGCATTCCCT

KENYATTA UNIVERSITY LIBRARY

CGAATGAATAATATAAGATTTTGATTATTACCTCCCTCTCTTACATTACTTTTAGTGAGCAGTATAGTAGAAAACGGAG
CTGGTACAGGTTGAACTGTTTATCCTCCCCTTTCATCAATTATCGCTCATGGTGGAGCCTCAGTTGATTTAGCTATTTTTT
CTCTACATTTAGCTGGTATTTCATCAATTTTAGGGGCCGTAAATTTCATTACTACAGTAATTAATATGCGATCAACAGGA
ATCACATTTGACCGGATACCTTTATTCGTTTGAGCTGTAGTATTGACAGCTCTTCTTTTACTTCTATCTCTACCTGTATTA
GCTGGAGCTATTACTATACTTTTAACAGACCGAAATTTAAACACCTCTTTCTTCGACCCGGCTGGTGGTGGAGACCCTAT
TTTATACCAACATTTATTT

>Boleae1

AACTTTATACTTCATTTTCGGGGCCTGAGCAGGAATAGTTTGGCACATCTCTTAGAATTTTAGTTCGAGCAGAACTAGGT
CATCCTGGAGCATTAATTGGAGATGATCAAATTTATAATGTAATTGTAACTGCTCACGCTTTCGTAATAATTTTTTTCAT
AGTTATACCTATTATAATCGGGGGGTTTGGAAATTGATTAGTCCCTTTAATATTAGGGGCACCAGATATAGCATTCCCC
CGAATAAATAATATAAGATTCTGGTTATTACCCCCTTCTCTCACTCTATTGTTAGTCAGCAGCATAGTGGAAAACGGAG
CCGGTACAGGTTGAACTGTATACCCACCATTATCATCTATTATTGCACATGGAGGAGCATCTGTCGATCTAGCCATCTTT
TCACTCCATTTAGCAGGTATCTCCTCAATCTTAGGAGCTGTAAATTTTATTACAACAGTAATTAATATGCGATCAACAGG
AATTACATTCGATCGAATACCTTTATTTGTTTGAGCAGTAGTACTAACAGCCCTACTACTTTTACTATCTCTACCTGTCTT
AGCCGGTGCTATCACAATACTTTTAACAGATCGAAATCTGAATACTTCCTTTTTCGACCCAGCAGGTGGAGGAGACCCA
ATCCTTTACCAACATTTATTT

> Boleae2

AACTTTATACTTCATTTTCGGAGCCTGAGCAGGAATAGTTTGGCACATCTCTTAGAATTTTAGTTCGAGCAGAACTAGGT
CATCCTGGAGCATTAATTGGAGATGATCAAATTTATAATGTAATTGTAACTGCTCACGCTTTCGTAATAATTTTTTTCAT
AGTTATACCTATTATAATCGGGGGGTTTGGAAATTGATTAGTCCCTTTAATATTAGGGGCACCAGATATAGCATTCCCC
CGAATAAATAATATAAGATTCTGGTTATTACCCCCTTCTCTCACTCTATTGTTAGTCAGCAGCATAGTGGAAAACGGAG
CCGGTACAGGTTGAACTGTATACCCACCATTATCATCTATTATTGCACATGGAGGAGCATCTGTCGATCTAGCCATCTTT
TCACTCCATTTAGCAGGTATCTCCTCAATCTTAGGAGCTGTAAATTTTATTACAACAGTAATTAATATGCGATCAACAGG
AATTACATTCGATCGAATACCTTTATTTGTTTGAGCAGTAGTACTAACAGCCCTACTACTTTTACTGTCTCTACCTGTCTT
AGCCGGTGCTATCACAATACTTTTAACAGATCGAAATCTGAATACTTCCTTTTTCGACCCAGCAGGTGGAGGAGACCCA
ATCCTTTACCAACATTTATTT

> Boleae3

AACTTTATACTTCATTTTCGGAGCCTGAGCAGGAATAGTTTGGCACATCTCTTAGAATTTTAGTTCGAGCAGAACTAGGT
CATCCTGGAGCATTAATTGGAGATGATCAAATTTATAATGTAATTGTAACTGCTCACGCTTTCGTAATAATTTTTTTCAT
AGTTATACCTATTATAATCGGGGGGTTTGGAAATTGATTAGTCCCTTTAATATTAGGGGCACCAGATATGGCATTCCCC
CGAATAAATAATATAAGATTCTGGTTATTACCCCCTTCTCTCACTCTATTGTTAGTCAGCAGCATAGTGGAAAACGGAG
CCGGTACAGGTTGAACTGTATACCCACCATTATCATCTATTATTGCACATGGAGGAGCATCTGTCGATCTAGCCATCTTT
TCACTCCATTTAGCAGGTATCTCCTCAATCTTAGGAGCTGTAAATTTTATTACAACAGTAATTAATATGCGATCAACAGG
AATTACATTCGATCGAATACCTTTATTTGTTTGAGCAGTAGTACTAACAGCCCTACTACTTTTACTATCTCTACCTGTCTT
AGCTGGTGCTATCACAATACTTTTAACAGATCGAAATCTGAATACTTCCTTTTTCGACCCAGCGGGTGGAGGAGACCCA
ATCCTTTACCAACATTTATTT

> Boleae4

AACTTTATACTTCATTTTCGGAGCCTGAGCAGGAATAGTTTGGCACATCTCTTAGAATTTTAGTTCGAGCAGAACTAGGT
CATCCTGGAGCATTAATTGGAGATGATCAAATTTATAATGTAATTGTAACTGCTCACGCTTTCGTAATAATTTTTTTCAT
AGTTATACCTATTATAATCGGGGGGTTTGGAAATTGATTAGTCCCTTTAATATTAGGGGCACCAGATATGGCATTCCCC
CGAATAAATAATATAAGATTCTGGTTATTACCCCCTTCTCTCACTCTATTGTTAGTCAGCAGCATAGTGGAAAACGGAG
CCGGTACAGGTTGAACTGTATACCCACCATTATCATCTATTATTGCACATGGAGGAGCATCTGTCGATCTAGCCATCTTT
TCACTCCATTTAGCAGGTATCTCCTCAATCTTAGGAGCTGTAAATTTTATTACAACAGTAATTAATATGCGATCAACAGG
AATTACATTCGATCGAATACCTTTATTTGTTTGAGCAGTAGTACTAACAGCCCTACTACTTTTACTATCTCTACCTGTCTT
AGCCGGTGCTATCACAATACTTTTAACAGATCGAAATCTGAATACTTCCTTTTTCGACCCAGCGGGTGGAGGAGACCCA
ATCCTTTACCAACATTTATTT

> Boleae5

AACTTTATACTTCATTTTCGGGGCCTGAGCAGGAATAGTTTGGCACATCTCTTAGAATTTTAGTTCGAGCAGAACTAGGT
CATCCTGGAGCATTAATTGGAGATGATCAAATTTATAATGTAATTGTAACTGCTCACGCTTTCGTAATAATTTTTTTCAT
AGTTATACCTATTATAATCGGGGGGTTTGGAAATTGATTAGTCCCTTTAATATTAGGGGCACCAGATATAGCATTCCCC
CGAATAAATAATATAAGATTCTGGTTATTACCCCCTTCTCTCACTCTATTGTTAGTCAGCAGCATAGTGGAAAACGGAG
CCGGTACAGGTTGAACTGTATACCCACCATTATCATCTATTATTGCACATGGAGGAGCATCTGTCGATCTAGCCATCTTT
TCACTCCATTTAGCAGGTATCTCCTCAATCTTAGGAGCTGTAAATTTTATTACAACAGTAATTAATATGCGATCAACAGG
AATTACATTCGATCGAATACCTTTATTTGTTTGAGCAGTAGTACTAACAGCCCTACTACTTTTACTATCTCTACCTGTCTT
AGCCGGTGCTATCACAATACTTTTAACAGATCGAAATCTGAATACTTCCTTTTTCGACCCAGCAGGTGGAGGAGACCCA
ATCCTTTACCAACATTTATTT

> Boleae6

AACTTTATACTTCATTTTCGGAGCCTGAGCAGGAATAGTTTGGCACATCTCTTAGAATTTTAGTTCGAGCAGAACTAGGT
CATCCTGGAGCATTAATTGGAGATGATCAAATTTATAATGTAATTGTAACTGCTCACGCTTTCGTAATAATTTTTTTCAT
AGTTATACCTATTATAATCGGGGGGTTTGGAAATTGATTAGTCCCTTTAATATTAGGGGCACCAGATATGGCATTCCCC
CGAATAAATAATATAAGATTCTGGTTATTACCCCCTTCTCTCACTCTATTGTTAGTCAGCAGCATAGTGGAAAACGGAG
CCGGTACAGGTTGAACTGTATACCCACCATTATCATCTATTATTGCACATGGAGGAGCATCTGTCGATCTAGCCATCTTT
TCACTCCATTTAGCAGGTATCTCCTCAATCTTAGGAGCTGTAAATTTTATTACAACAGTAATTAATATACGATCAACAGG
AATTACATTCGATCGAATACCTTTATTTGTTTGAGCAGTAGTACTAACAGCCCTACTACTTTTACTATCTCTACCTGTCTT
AGCCGGTGCTATCACAATACTTTTAACAGATCGAAATCTGAATACTTCCTTTTTCGACCCAGCAGGTGGAGGAGACCCA
ATCCTTTACCAACATTTATTT

> Boleae7

AACTTTATACTTCATTTTCGGAGCCTGAGCAGGAATAGTTTGGCACATCTCTTAGAATTTTAGTTCGAGCAGAACTAGGT
CATCCTGGAGCATTAATTGGAGATGATCAAATTTATAATGTAATTGTAACTGCTCACGCTTTCGTAATAATTTTTTTCAT
AGTTATACCTATTATAATCGGGGGGTTTGGAAATTGATTAGTCCCTTTAATATTAGGGGCACCAGATATGGCATTCCCC

CGAATAAATAATATAAGATTCTGGTTATTACCCCCTTCTCTCACTCTATTGTTAGTCAGCAGCATAGTGGAAAACGGAG
CCGGTACAGGTTGAACTGTATACCCACCATTATCATCTATTATTGCACATGGAGGAGCATCTGTCGATCTAGCCATCTTT
TCACTCCATTTAGCAGGTATCTCCTCAATCTTAGGAGCTGTAAATTTTATTACAACAGTAATTAATATGCGATCAACAGG
AATTACATTCGATCGAATACCTTTATTTGTTTGAGCAGTAGTACTAACAGCCCTACTACTTTTACTATCTCTACCTGTCTT
AGCCGGTGCTATCACAATACTTTTAACAGATCGAAATCTGAATACTTCCTTTTTCGACCCAGCGGGTGGAGGAGACCCA
ATCCTTTACCAACATTTATTT
> Boleae8
AACTTTATACTTCATTTTCGGAGCCTGAGCAGGAATAGTTTGGCACATCTCTTAGAATTTTAGTTCGAGCAGAACTAGGT
CATCCTGGAGCATTAATTGGAGATGATCAAATTTATAATGTAATTGTAACTGCTCACGCTTTCGTAATAATTTTTTTCAT
AGTTATACCTATTATAATCGGGGGGTTTGGAAATTGATTAGTCCCTTTAATATTAGGGGCACCAGATATGGCATTCCCC
CGAATAAATAATATAAGATTCTGGTTATTACCCCCTTCTCTCACTCTATTGTTAGTCAGCAGCATAGTGGAAAACGGAG
CCGGTACAGGTTGAACTGTATACCCACCATTATCATCTATTATTGCACATGGAGGAGCATCTGTCGATCTAGCCATCTTT
TCACTCCATTTAGCAGGTATCTCCTCAATCTTAGGAGCTGTAAATTTTATTACAACAGTAATTAATATGCGATCAACAGG
AATTACATTCGATCGAATACCTTTATTTGTTTGAGCAGTAGTACTAACAGCCCTACTACTTTTACTATCTCTACCTGTCTT
AGCTGGTGCTATCACAATACTTTTAACAGATCGAAATCTGAATACTTCCTTTTTCGACCCAGCGGGTGGAGGAGACCCA
ATCCTTTACCAACATTTATTT
> Boleae9
AACTTTATACTTCATTTTCGGGGCCTGAGCAGGAATAGTTTGGCACATCTCTTAGAATTTTAGTTCGAGCAGAACTAGGT
CATCCTGGAGCATTAATTGGAGATGATCAAATTTATAATGTAATTGTAACTGCTCACGCTTTCGTAATAATTTTTTTCAT
AGTTATACCTATTATAATCGGGGGGTTTGGAAATTGATTAGTCCCTTTAATATTAGGGGCACCAGATATAGCATTCCCC
CGAATAAATAATATAAGATTCTGGTTATTACCCCCTTCTCTCACTCTATTGTTAGTCAGCAGCATAGTGGAAAACGGAG
CCGGTACAGGTTGAACTGTATACCCACCATTATCATCTATTATTGCACATGGAGGAGCATCTGTCGATCTAGCCATCTTT
TCACTCCATTTAGCAGGTATCTCCTCAATCTTAGGAGCTGTAAATTTTATTACAACAGTAATTAATATGCGATCAACAGG
AATTACATTCGATCGAATACCTTTATTTGTTTGAGCAGTAGTACTAACAGCCCTACTACTTTTACTATCTCTACCTGTCTT
AGCCGGTGCTATCACAATACTTTTAACAGATCGAAATCTGAATACTTCCTTTTTCGACCCAGCAGGTGGAGGAGACCCA
ATCCTTTACCAACATTTATTT
> Boleae10
AACTTTATACTTCATTTTCGGGGCCTGAGCAGGAATAGTTTGGCACATCTCTTAGAATTTTAGTTCGAGCAGAACTAGGT
CATCCTGGAGCATTAATTGGAGATGATCAAATTTATAATGTAATTGTAACTGCTCACGCTTTCGTAATAATTTTTTTCAT
AGTTATACCTATTATAATCGGGGGGTTTGGAAATTGATTAGTCCCTTTAATATTAGGGGCACCAGATATAGCATTCCCC
CGAATAAATAATATAAGATTCTGGTTATTACCCCCTTCTCTCACTCTACTGTTAGTCAGCAGCATAGTGGAAAACGGAG
CCGGTACAGGTTGAACTGTATACCCACCATTATCATCTATTATTGCACATGGAGGAGCATCTGTCGATCTAGCCATCTTT
TCACTCCATTTAGCAGGTATCTCCTCAATCTTAGGAGCTGTAAATTTTATTACAACAGTAATTAATATGCGATCAACAGG
AATTACATTCGATCGAATACCTTTATTTGTTTGAGCAGTAGTACTAACAGCCCTACTACTTTTACTATCTCTACCTGTCTT
AGCCGGTGCTATCACAATACTTTTAACAGATCGAAATCTGAATACTTCCTTTTTCGACCCAGCAGGTGGAGGAGACCCA
ATCCTTTACCAACATTTATTT
>BK1
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTTGGAACATCTCTTAGAATTTTAGTTCGAGCTGAACTGGGTC
ATCCAGGAGCGTTAATTGGAGACGACCAAATTTATAATGTAATTGTAACAGCCCATGCCTTCGTAATAATTTTCTTTATA
GTTATACCAATTATAATTGGTGGATTCGGAAATTGACTTGTTCCCCTAATATTAGGTGCCCCTGATATAGCTTTTCCACG
AATAAACAATATAAGATTTTGATTATTACCTCCTTCCCTTACACTACTATTAGTGAGAAGTATAGTAGAAAATGGAGCT
GGTACAGGTTGAACAGTTTACCCACCCCTATCATCGGTTATTGCACACGGAGGAGCTTCAGTTGATCTAGCTATTTTCTC
ACTCCACTTAGCTGGTATCTCTTCAATTTTAGGAGCAGTTAATTTCATTACAACAGTAATTAATATACGATCAACAGGA
ATCTCTTTTGATCGAATACCTCTTTTCGTATGAGCAGTTGTATTAACAGCCTTATTACTTTTATTATCATTACCAGTTTTA
GCAGGAGCTATTACTATATTATTAACAGATCGAAACTTAAATACTTCCTTTTTTGACCCCGCCGGAGGGGGAGACCCTA
TTCTTTACCAACATTTATTT
>BK2
AACTTTATATTTTATCTTCGGGGCCTGAGCAGGAATAGTTGGAACATCTCTTAGAATTTTAGTTCGAGCTGAACTGGGTC
ATCCAGGAGCATTAATTGGAGACGACCAAATTTATAATGTAATTGTAACAGCCCATGCCTTCGTAATAATTTTCTTTATA
GTTATACCAATTATAATTGGTGGATTCGGAAATTGACTTGTTCCCCTAATATTAGGTGCTCCTGATATAGCCTTTCCACG
AATAAACAATATAAGATTTTGATTATTACCTCCTTCCCTTACACTACTATTAGTGAGAAGCATAGTAGAAAATGGAGCT
GGTACAGGTTGAACAGTTTACCCACCCCTATCATCGGTTATTGCACACGGAGGAGCTTCAGTTGATCTAGCTATTTTCTC
ACTCCACTTAGCTGGTATCTCTTCAATTTTAGGGGCAGTTAATTTCATTACAACAGTAATTAATATACGATCAACAGGA
ATCTCTTTTGATCGAATACCTCTTTTCGTGTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTTA
GCAGGAGCTATTACTATATTATTAACAGATCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGGGGAGACCCTA
TTCTTTACCAACATTTATTT
>BK3
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTTGGAACATCTCTTAGAATTTTAGTTCGAGCTGAACTGGGTC
ATCCAGGAGCGTTAATTGGAGACGACCAAATTTATAATGTAATTGTAACAGCCCATGCCTTCGTAATAATTTTCTTTATA
GTTATACCAATTATAATTGGTGGATTCGGAAATTGACTTGTTCCCCTAATATTAGGTGCTCCTGATATAGCTTTTCCACG
AATAAACAATATAAGATTTTGATTATTACCTCCTTCCCTTACACTACTATTAGTGAGAAGTATAGTAGAAAATGGAGCT
GGTACAGGTTGAACAGTTTACCCACCCCTATCATCGGTTATTGCACACGGAGGAGCTTCAGTTGATCTAGCTATTTTCTC
ACTCCACTTAGCCGGTATCTCTTCAATTTTAGGAGCAGTTAATTTCATTACAACAGTAATTAATATGCGATCAACAGGA
ATCTCTTTTGATCGAATACCTCTTTTCGTATGAGCAGTTGTATTAACAGCCTTATTACTTTTATTATCATTACCAGTTTTA
GCAGGAGCTATTACTATATTATTAACAGATCGAAACTTAAATACTTCCTTTTTTGACCCCGCCGGAGGGGGAGACCCTA
TTCTTTACCAACATTTATTT
>BK4
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTTGGAACATCTCTTAGAATTTTAGTTCGAGCTGAACTGGGTC
ATCCAGGAGCATTAATTGGAGACGACCAAATTTATAATGTAATTGTAACAGCCCATGCCTTCGTAATAATTTTCTTTATA
GTTATACCAATTATAATTGGTGGATTCGGAAATTGACTTGTTCCCCTAATATTAGGTGCTCCTGATATAGCTTTTCCACG

AATAAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACACTACTATTAGTGAGAAGCATAGTAGAAAATGGAGCT
GGTACAGGTTGAACAGTTTACCCACCCCTATCATCGGTTATTGCACACGGAGGAGCTTCAGTTGATCTAGCTATTTTCTC
ACTCCACTTAGCCGGTATCTCTTCAATTTTAGGGGCAGTTAATTTCATTACAACAGTAATTAATATACGATCAACAGGA
ATCTCTTTTGATCGAATACCTCTTTTCGTATGAGCAGTTGTATTAACAGCTTTATTACTTTTATTGTCATTACCAGTTTTA
GCAGGAGCTATTACTATATTATTAACAGATCGAAACTTAAATACTTCCTTTTTTGACCCCGCCGGAGGGGGAGACCCTA
TTCTTTACCAACATTTATTT
>BK5
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTTGGAACATCTCTTAGAATTTTAGTTCGAGCTGAACTGGGTC
ATCCAGGAGCGTTAATTGGAGACGACCAAATTTATAATGTAATTGTAACAGCCCATGCCTTCGTAATAATTTTCTTTATG
GTTATACCAATTATAATTGGTGGGTTCGGAAATTGACTTGTTCCCCTAATATTAGGTGCTCCTGATATAGCTTTTCCACG
AATAAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACACTACTATTAGTGAGAAGTATAGTAGAAAATGGAGCT
GGTACAGGTTGAACAGTTTACCCACCCCTATCATCAGTTATTGCACACGGAGGAGCTTCAGTTGATCTAGCTATTTTCTC
ACTCCACTTAGCCGGTATCTCTTCAATTTTAGGGGCAGTTAATTTCATTACAACAGTAATTAATATACGATCAACAGGA
ATCTCTTTTGATCGAATACCTCTTTTCGTATGAGCAGTTGTATTAACAGCCTTATTACTTTTATTATCATTACCAGTTTTA
GCAGGAGCTATTACTATATTATTAACAGATCGAAACTTAAATACTTCCTTTTTTGACCCCGCCGGAGGGGGAGACCCTA
TTCTTTACCAACATTTATTT
>BK6
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTTGGAACATCTCTTAGAATTTTAGTTCGAGCTGAACTGGGTC
ATCCAGGAGCGTTAATTGGAGACGACCAAATTTATAATGTAATTGTAACAGCCCATGCCTTCGTAATAATTTTCTTTATA
GTTATACCAATTATAATTGGTGGATTCGGAAATTGACTTGTTCCCCTAATATTAGGTGCTCCTGATATAGCTTTTCCACG
AATAAACAATATAAGATTTTGATTATTACCTCCTTCCCTTACACTACTATTAGTGAGAAGTATAGTAGAAAATGGAGCT
GGTACAGGTTGAACAGTTTACCCACCCCTATCATCGGTTATTGCACACGGAGGAGCTTCAGTTGATCTAGCTATTTTCTC
ACTCCACTTAGCCGGTATCTCTTCAATTTTAGGAGCAGTTAATTTCATTACAACAGTAATTAATATACGATCAACAGGA
ATCTCTTTTGATCGAATACCTCTTTTCGTATGAGCAGTTGTATTAACAGCCCTATTACTTTTATTATCATTACCAGTTTTA
GCAGGAGCTATTACTATATTATTAACAGATCGAAACTTAAATACTTCCTTTTTTGACCCCGCCGGAGGGGGAGACCCTA
TTCTTTACCAACATTTATTT
>BK7
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTTGGAACATCTCTTAGAATTTTAGTTCGAGCTGAACTGGGTC
ATCCAGGAGCATTAATTGGAGACGACCAAATTTATAATGTAATTGTAACAGCCCATGCCTTCGTAATAATTTTCTTTATA
GTTATACCAATTATAATTGGTGGATTCGGAAATTGACTTGTTCCCCTAATATTAGGTGCTCCTGATATAGCTTTTCCACG
AATAAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACACTACTATTAGTGAGAAGCATAGTAGAAAATGGAGCT
GGTACAGGTTGAACAGTTTACCCACCCCTATCATCGGTTATTGCACACGGAGGAGCTTCAGTTGATCTAGCTATTTTCTC
ACTCCACTTAGCCGGTATCTCTTCAATTTTAGGGGCAGTTAATTTCATTACAACAGTAATTAATATACGATCAACAGGA
ATCTCTTTTGATCGAATACCTCTTTTCGTATGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTTA
GCAGGAGCTATTACTATATTATTAACAGATCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGGGGAGACCCTA
TTCTTTACCAACATTTATTT
>BK8
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTTGGAACATCTCTTAGAATTTTAGTTCGAGCTGAACTGGGTC
ATCCAGGAGCGTTAATTGGAGACGACCAAATTTATAATGTAATTGTAACAGCCCATGCCTTCGTAATAATTTTCTTTATA
GTTATACCAATTATAATTGGTGGATTCGGAAATTGACTTGTTCCCCTAATATTAGGTGCTCCTGATATAGCTTTTCCACG
AATAAACAATATAAGATTTTGATTATTACCTCCTTCCCTTACACTACTATTAGTGAGAAGTATAGTAGAAAATGGAGCT
GGTACAGGTTGAACAGTTTACCCACCCCTATCATCGGTTATTGCACACGGAGGAGCTTCAGTTGATCTAGCTATTTTCTC
ACTCCACTTAGCCGGTATCTCTTCAATTTTAGGAGCAGTTAATTTCATTACAACAGTAATTAATATACGATCAACAGGA
ATCTCTTTTGATCGAATACCTCTTTTCGTATGAGCAGTTGTATTAACAGCCTTATTACTTTTATTATCATTACCAGTTTTA
GCAGGAGCTATTACTATATTATTAACAGATCGAAACTTAAATACTTCCTTTTTTGACCCCGCCGGAGGGGGAGACCCTA
TTCTTTACCAACATTTATTT
>BK9
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTTGGAACATCTCTTAGAATTTTAGTTCGAGCTGAACTGGGTC
ATCCAGGAGCGTTAATTGGAGACGACCAAATTTATAATGTAATTGTAACAGCCCATGCCTTCGTAATAATTTTCTTTATG
GTTATACCAATTATAATTGGTGGGTTCGGAAATTGACTTGTTCCCCTAATATTAGGTGCTCCTGATATAGCTTTTCCACG
AATAAACAATATAAGATTTTGATTATTACCTCCTTCCCTTACACTACTATTAGTGAGAAGTATAGTAGAAAATGGAGCT
GGTACAGGTTGAACAGTTTACCCACCCCTATCATCGGTTATTGCACACGGAGGAGCTTCAGTTGATCTAGCTATTTTCTC
ACTCCACTTAGCCGGTATCTCTTCAATTTTAGGAGCAGTTAATTTCATTACAACAGTAATTAATATACGATCAACAGGA
ATCTCTTTTGATCGAATACCTCTTTTCGTATGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTTA
GCAGGAGCTATTACTATATTATTAACAGATCGAAACTTAAATACTTCCTTTTTTGACCCCGCCGGAGGGGGAGACCCTA
TTCTTTACCAACATTTATTT
>Bp1
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATGCGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATACTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>Bp2
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC

GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATGCGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATACTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCCGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT

>Bp3

AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATGCGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATACTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCCGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT

>Bp4

AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAAC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATGCGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATACTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT

>Bp5

AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATGCGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATACTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCCGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT

>Bp6

AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATGCGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATACTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCCGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT

>Bp7

AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATGCGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATACTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCCGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT

>Bp8

AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATGCGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATACTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT

>Bp9

AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC

GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATGCGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATACTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>Bp10
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGACCTAGCTATTTTTT
CACTTCACTTAGCGGGTATGTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATGCGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGGGCTATTACTATACTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGAGGAGATCCT
ATTCTTTACCAACATTTATTT
>Bz1
AACCTTATATTTTATTTTCGGAGCCTGAGCAGGTATAGTTTGGAACATCCCTTAGAATTTTAGTTCGTGCTGAGCTAGGA
CATCCCGGAGCACTAATTGGAGATGATCAGATTTATAATGTAATTGTAACAGCACACGCTTTCGTAATAATTTTCTTTAT
AGTAATACCTATTATAATTGGAGGATTTGGAAATTGACTTGTTCCCCTAATATTAGGAGCACCCGACATAGCATTCCCA
CGAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACGCTGCTATTAGTGAGAAGTATAGTAGAAAACGGAG
CTGGTACAGGTTGAACAGTTTATCCTCCCCTATCATCTGTTATTGCTCACGGAGGAGCTTCAGTTGATCTAGCTATTTTC
TCACTTCACTTAGCTGGTATTTCCTCAATTCTAGGAGCAGTTAATTTTATTACAACTGTTATTAACATACGTTCAACAGG
AATTTCATTTGATCGAATACCTCTCTTCGTTTGAGCAGTTGTATTAACAGCCCTATTACTTCTATTATCATTGCCAGTTTT
AGCAGGAGCTATTACTATATTATTAACAGATCGAAATTTAAATACTTCATTTTTTGACCCTGCGGGAGGAGGAGACCCT
ATTCTCTATCAACACTTATTT
>Bz2
AACCTTATATTTTATTTTCGGAGCCTGAGCAGGTATAGTTTGGAACATCCCTTAGAATTTTAGTTCGTGCTGAGCTAGGA
CATCCCGGAGCACTAATTGGAGATGATCAGATTTATAATGTAATTGTAACAGCACACGCTTTCGTAATAATTTTCTTTAT
AGTAATACCTATTATAATTGGAGGATTTGGAAATTGACTTGTTCCCCTAATATTAGGAGCACCCGACATAGCATTCCCA
CGAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACGCTGCTATTAGTGAGAAGTATAGTAGAAAACGGAG
CTGGTACAGGTTGAACAGTTTATCCTCCCCTATCATCTGTTATTGCTCACGGAGGAGCTTCAGTTGATCTAGCTATTTTC
TCACTTCACTTAGCTGGTATTTCCTCAATTCTAGGAGCAGTTAATTTTATTACAACTGTTATTAACATACGTTCAACAGG
AATTTCATTTGATCGAATACCTCTCTTCGTTTGAGCAGTTGTATTAACAGCCCTATTACTTCTATTATCATTGCCAGTTTT
AGCAGGAGCTATTACTATATTATTAACAGATCGAAATTTAAATACTTCATTTTTTGACCCTGCGGGAGGAGGAGACCCT
ATTCTCTATCAACACTTATTT
>Bz3
AACCTTATATTTTATTTTCGGAGCCTGAGCAGGTATAGTTTGGAACATCCCTTAGAATTTTAGTTCGTGCTGAGCTAGGA
CATCCCGGAGCACTAATTGGAGATGATCAGATTTATAATGTAATTGTAACAGCACACGCTTTCGTAATAATTTTCTTTAT
AGTAATACCTATTATAATTGGAGGATTTGGAAATTGACTTGTTCCCCTAATATTAGGAGCACCCGACATAGCATTCCCA
CGAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACGCTGCTATTAGTGAGAAGTATAGTAGAAAACGGAG
CTGGTACAGGTTGAACAGTTTATCCTCCCCTATCATCTGTTATTGCTCACGGAGGAGCTTCAGTTGATCTAGCTATTTTC
TCACTTCACTTAGCTGGTATTTCCTCAATTCTAGGAGCAGTTAATTTTATTACAACTGTTATTAACATACGTTCAACAGG
AATTTCATTTGATCGAATACCTCTCTTCGTTTGAGCAGTTGTATTAACAGCCCTATTACTTCTATTATCATTGCCAGTTTT
AGCAGGAGCTATTACTATATTATTAACAGATCGAAATTTAAATACTTCATTTTTTGACCCTGCGGGAGGAGGAGACCCT
ATTCTCTATCAACACTTATTT
>Bz4
AACCTTATATTTTATTTTCGGAGCCTGAGCAGGTATAGTTTGGAACATCCCTTAGAATTTTAGTTCGTGCTGAGCTAGGA
CATCCCGGAGCACTAATTGGAGATGATCAGATTTATAATGTAATTGTAACAGCACACGCTTTCGTAATAATTTTCTTTAT
AGTAATACCTATTATAATTGGAGGATTTGGAAATTGACTTGTTCCCCTAATATTAGGAGCACCCGACATAGCATTCCCA
CGAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACGCTGCTATTAGTGAGAAGTATAGTAGAAAACGGAG
CTGGTACAGGTTGAACAGTTTATCCTCCCCTATCATCTGTTATTGCTCACGGAGGAGCTTCAGTTGATCTAGCTATTTTC
TCACTTCACTTAGCTGGTATTTCCTCAATTCTAGGAGCAGTTAATTTTATTACAACTGTTATTAACATACGTTCAACAGG
AATTTCATTTGATCGAATACCTCTCTTCGTTTGAGCAGTTGTATTAACAGCCCTATTACTTCTATTATCATTGCCAGTTTT
AGCAGGAGCTATTACTATATTATTAACAGATCGAAATTTAAATACTTCATTTTTTGACCCTGCGGGAGGAGGAGACCCT
ATTCTCTATCAACACTTATTT
>Bz5
AACCTTATATTTTATTTTCGGAGCCTGAGCAGGTATAGTTTGGAACATCTCTTAGAATTTTAGTTCGTGCTGAACTAGGA
CACCCCGGAGCACTAATTGGAGATGATCAGATTTATAATGTAATTGTAACAGCACACGCTTTCGTAATAATTTTCTTTAT
AGTAATACCTATTATAATTGGAGGATTTGGAAATTGACTTGTTCCCCTAATATTAGGAGCACCCGACATAGCATTCCCA
CGAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACGCTGCTATTAGTGAGAAGTATAGTAGAAAACGGAG
CTGGTACAGGTTGAACAGTTTATCCTCCCCTATCATCTGTTATTGCTCACGGAGGAGCTTCAGTTGATCTAGCTATTTTC
TCACTTCACTTAGCTGGTATTTCTTCAATTCTAGGAGCAGTTAATTTTATTACAACTGTTATTAACATACGTTCAACAGG
AATTTCATTTGATCGAATACCTCTCTTCGTTTGAGCAGTTGTATTAACAGCCCTATTACTTCTATTATCATTACCAGTTTT
AGCGGGGGCTATTACTATATTATTAACAGATCGAAATTTAAATACTTCATTTTTTGACCCTGCAGGAGGAGGAGACCCT
ATTCTCTATCAACACTTATTT
>Bz6
AACCTTATATTTTATTTTCGGAGCCTGAGCAGGTATAGTTTGGAACATCTCTTAGAATTTTAGTTCGTGCTGAACTAGGA
CACCCCGGAGCACTAATTGGAGATGATCAGATTTATAATGTAATTGTAACAGCACACGCTTTCGTAATAATTTTCTTTAT
AGTAATACCTATTATAATTGGAGGATTTGGAAATTGACTTGTTCCCCTAATATTAGGAGCACCCGACATAGCATTCCCA

CGAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACGCTGCTATTAGTGAGAAGTATAGTAGAAAACGGAG
CTGGTACAGGTTGAACAGTTTATCCTCCCCTATCATCTGTTATTGCTCACGGAGGAGCTTCAGTTGATCTAGCTATTTTC
TCACTTCACTTAGCTGGTATTTCTTCAATTCTAGGAGCAGTTAATTTTATTACAACTGTTATTAACATACGTTCAACAGG
AATTTCATTTGATCGAATACCTCTCTTCGTTTGAGCAGTTGTATTAACAGCCCTATTACTTCTATTATCATTACCAGTTTT
AGCGGGGGCTATTACTATATTATTAACAGATCGAAATTTAAATACTTCATTTTTTGACCCTGCAGGAGGAGGAGACCCT
ATTCTCTATCAACACTTATTT
>Bz7
AACCTTATATTTTATTTTCGGAGCCTGAGCAGGTATAGTTTGGAACATCTCTTAGAATTTTAGTTCGTGCTGAACTAGGA
CACCCCGGAGCACTAATTGGAGATGATCAGATTTATAATGTAATTGTAACAGCACACGCTTTCGTAATAATTTTCTTTAT
AGTAATACCTATTATAATTGGAGGATTTGGAAATTGACTTGTTCCCCTAATATTAGGAGCACCCGACATAGCATTCCCA
CGAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACGCTGCTATTAGTGAGAAGTATAGTAGAAAACGGAG
CTGGTACAGGTTGAACAGTTTATCCTCCCCTATCATCTGTTATTGCTCACGGAGGAGCTTCAGTTGATCTAGCTATTTTC
TCACTTCACTTAGCTGGTATTTCTTCAATTCTAGGAGCAGTTAATTTTATTACAACTGTTATTAACATACGTTCAACAGG
AATTTCATTTGATCGAATACCTCTCTTCGTTTGAGCAGTTGTATTAACAGCCCTATTACTTCTATTATCATTACCAGTTTT
AGCGGGGGCTATTACTATATTATTAACAGATCGAAATTTAAATACTTCATTTTTTGACCCTGCAGGAGGAGGAGACCCT
ATTCTCTATCAACACTTATTT
>Bz8
AACCTTATATTTTATTTTCGGAGCCTGAGCAGGTATAGTTTGGAACATCTCTTAGAATTTTAGTTCGTGCTGAACTAGGA
CACCCCGGAGCACTAATTGGAGATGATCAGATTTATAATGTAATTGTAACAGCACACGCTTTCGTAATAATTTTCTTTAT
AGTAATACCTATTATAATTGGAGGATTTGGAAATTGACTTGTTCCCCTAATATTAGGAGCACCCGACATAGCATTCCCA
CGAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACGCTGCTATTAGTGAGAAGTATAGTAGAAAACGGAG
CTGGTACAGGTTGAACAGTTTATCCTCCCCTATCATCTGTTATTGCTCACGGAGGAGCTTCAGTTGATCTAGCTATTTTC
TCACTTCACTTAGCTGGTATTTCTTCAATTCTAGGAGCAGTTAATTTTATTACAACTGTTATTAACATACGTTCAACAGG
AATTTCATTTGATCGAATACCTCTCTTCGTTTGAGCAGTTGTATTAACAGCCCTATTACTTCTATTATCATTACCAGTTTT
AGCGGGGGCTATTACTATATTATTAACAGATCGAAATTTAAATACTTCATTTTTTGACCCTGCAGGAGGAGGAGACCCT
ATTCTCTATCAACACTTATTT
>Bz9
AACCTTATATTTTATTTTCGGAGCCTGAGCAGGTATAGTTTGGAACATCTCTTAGAATTTTAGTTCGTGCTGAACTAGGA
CACCCCGGAGCACTAATTGGAGATGATCAGATTTATAATGTAATTGTAACAGCACACGCTTTCGTAATAATTTTCTTTAT
AGTAATACCTATTATAATTGGAGGATTTGGAAATTGACTTGTTCCCCTAATATTAGGAGCACCCGACATAGCATTCCCA
CGAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACGCTGCTATTAGTGAGAAGTATAGTAGAAAACGGAG
CTGGTACAGGTTGAACAGTTTATCCTCCCCTATCATCTGTTATTGCTCACGGAGGAGCTTCAGTTGATCTAGCTATTTTC
TCACTTCACTTAGCTGGTATTTCTTCAATTCTAGGAGCAGTTAATTTTATTACAACTGTTATTAACATACGTTCAACAGG
AATTTCATTTGATCGAATACCTCTCTTCGTTTGAGCAGTTGTATTAACAGCCCTATTACTTCTATTATCATTACCAGTTTT
AGCGGGGGCTATTACTATATTATTAACAGATCGAAATTTAAATACTTCATTTTTTGACCCTGCAGGAGGAGGAGACCCT
ATTCTCTATCAACACTTATTT
>Bz10
AACCTTATATTTTATTTTCGGAGCCTGAGCAGGTATAGTTTGGAACATCCCTTAGAATTTTAGTTCGTGCTGAGCTAGGA
CATCCCGGAGCACTAATTGGAGATGATCAGATTTATAATGTAATTGTAACAGCACACGCTTTCGTAATAATTTTCTTTAT
AGTAATACCTATTATAATTGGAGGATTTGGAAATTGACTTGTTCCCCTAATATTAGGAGCACCCGACATAGCATTCCCA
CGAATGAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACGCTGCTATTAGTGAGAAGTATAGTAGAAAACGGAG
CTGGTACAGGTTGAACAGTTTATCCTCCCCTATCATCTGTTATTGCTCACGGAGGAGCTTCAGTTGATCTAGCTATTTTC
TCACTTCACTTAGCTGGTATTTCCTCAATTCTAGGAGCAGTTAATTTTATTACAACTGTTATTAACATACGTTCAACAGG
AATTTCATTTGATCGAATACCTCTCTTCGTTTGAGCAGTTGTATTAACAGCCCTATTACTTCTATTATCATTGCCAGTTTT
AGCAGGAGCTATTACTATATTATTAACAGATCGAAATTTAAATACTTCATTTTTTGACCCTGCGGGAGGAGGAGACCCT
ATTCTCTATCAACACTTATTT
>Bdorsalis1
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATAAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGATCTAGCTATTTTTT
CACTTCACTTAGCAGGTATTTCCTCAATTTTAGGAGCAGTTAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGAGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGGGGAGACCCT
ATTCTTTACCAACATTTATTT
>Bdorsalis2
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATAAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGATCTAGCTATTTTTT
CACTTCACTTAGCAGGTATTTCCTCAATTTTAGGAGCAGTTAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGAGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGGGGAGACCCT
ATTCTTTACCAACATTTATTT
>Bdorsalis3
AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC

GAATAAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGATCTAGCTATTTTTT
CACTTCACTTAGCAGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCGGGAGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGGGGAGACCCT
ATTCTTTACCAACATTTATTT

>Bdorsalis4

AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATAAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGATCTAGCTATTTTTT
CACTTCACTTAGCAGGTATTTCCTCAATTTTAGGAGCAGTTAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGAGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGGGGAGACCCT
ATTCTTTACCAACATTTATTT

>Bdorsalis5

AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATAAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGATCTAGCTATTTTTT
CACTTCACTTAGCAGGTATTTCCTCAATTTTAGGAGCAGTTAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGAGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGGGGAGACCCT
ATTCTTTACCAACATTTATTT

>Bdorsalis7

AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATAAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGATCTAGCTATTTTTT
CACTTCACTTAGCAGGTATTTCCTCAATTTTAGGAGCAGTTAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGAGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGGGGAGACCCT
ATTCTTTACCAACATTTATTT

>Bdorsalis8

AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATAAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGATCTAGCTATTTTTT
CACTTCACTTAGCAGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCGGGAGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGGGGAGACCCT
ATTCTTTACCAACATTTATTT

>Bdorsalis9

AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATAAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGATCTAGCTATTTTTT
CACTTCACTTAGCAGGTATTTCCTCAATTTTAGGAGCAGTTAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCAGGAGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGGGGAGACCCT
ATTCTTTACCAACATTTATTT

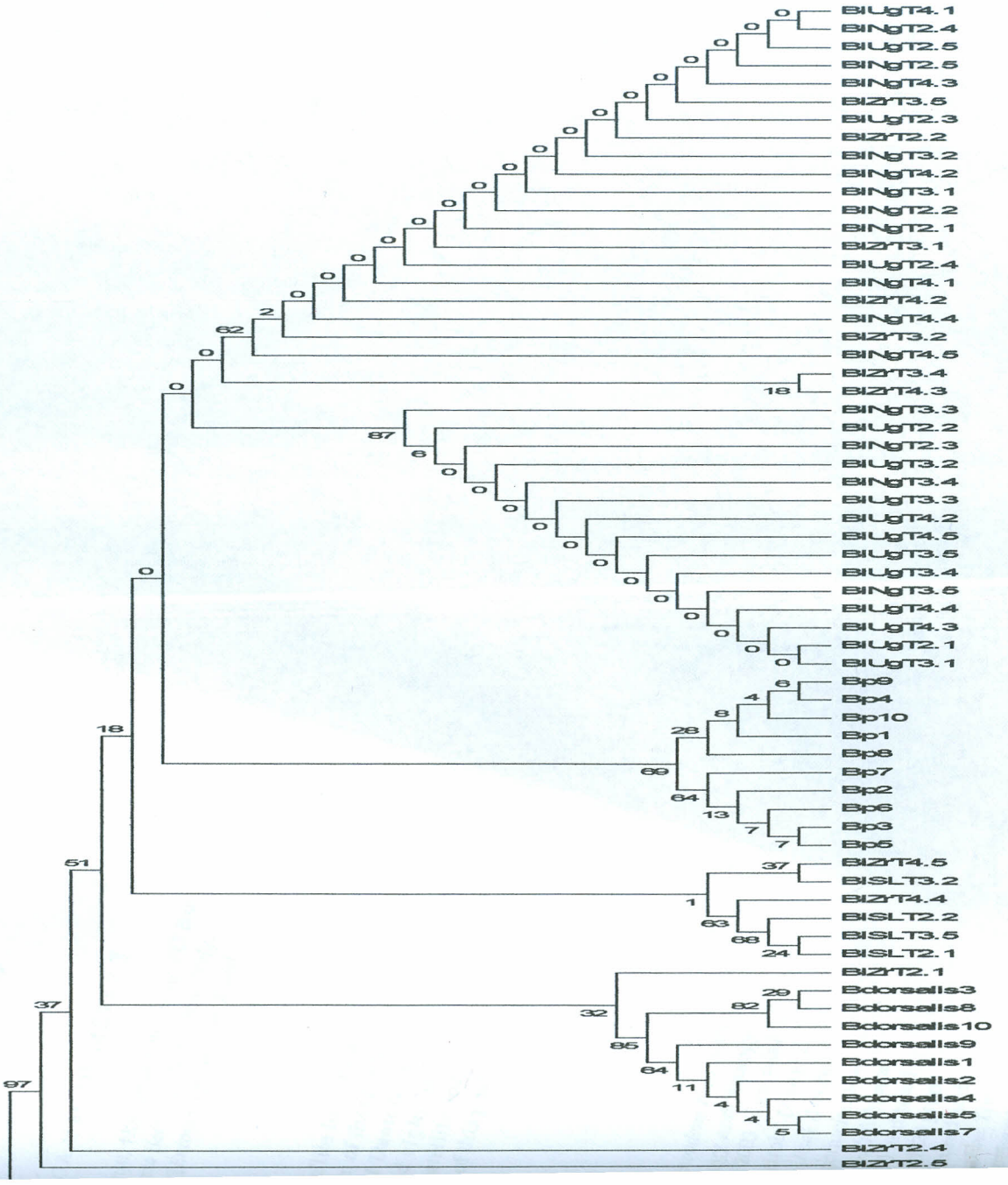
>Bdorsalis10

AACTTTATATTTTATCTTCGGAGCCTGAGCAGGAATAGTAGGAACATCCCTTAGAATTTTAGTCCGAGCTGAACTCGGT
CACCCAGGAGCTTTAATCGGTGACGATCAAATTTATAATGTAATTGTAACAGCTCATGCTTTCGTAATAATTTTCTTTAT
AGTTATACCAATTATAATTGGTGGATTTGGAAATTGACTTGTTCCTTTAATATTAGGAGCTCCCGATATAGCATTTCCAC
GAATAAATAATATAAGATTTTGATTATTACCTCCTTCCCTTACATTACTATTAGTAAGAAGTATAGTAGAAAACGGAGC
TGGTACAGGTTGAACAGTTTACCCACCCCTATCATCTGTTATTGCACACGGAGGAGCTTCAGTTGATCTAGCTATTTTTT
CACTTCACTTAGCAGGTATTTCCTCAATTTTAGGAGCAGTAAATTTCATTACAACAGTAATTAATATACGATCGACAGG
AATCACCTTTGATCGAATACCTCTATTCGTTTGAGCAGTTGTATTAACAGCTTTATTACTTTTATTATCATTACCAGTTTT
AGCGGGAGCTATTACTATATTACTAACAGACCGAAACTTAAATACTTCCTTTTTTGACCCTGCCGGAGGGGGAGACCCT
ATTCTTTACCAACATTTATTT

KENYATTA UNIVERSITY LIBRARY

APPENDIX VI

Maximum Parsimony phylogenetic tree generated by MEGA4 (Tamura *et al.*, 2007) for COI data.



KENYATTA ... LIBRARY

Molecular Ecology Resources (2008) 8, 1509–1511 doi: 10.1111/j.1755-0998.2008.02335.x

PERMANENT GENETIC RESOURCES

Isolation and characterization of microsatellite markers in the newly discovered invasive fruit fly pest in Africa, *Bactrocera invadens* (Diptera: Tephritidae)

F. KHAMIS,*†‡ N. KARAM,‡ C. R. GUGLIELMINO,§ S. EKESI,* D. MASIGA,*† M. DE MEYER,¶ E. U. KENYA† and A. R. MALACRIDA‡

*International Centre of Insect Physiology and Ecology, PO Box 30772-00100 GPO, Nairobi, Kenya, †Kenyatta University, PO Box 43844-00100 GPO, Nairobi, Kenya, ‡Dipartimento di Biologia Animale, Università di Pavia, Piazza Botta 9, 27100 Pavia, Italy, §Dipartimento di Genetica e Microbiologia, Università di Pavia, via Ferrata 1, 27100 Pavia, Italy, ¶Entomology Section, Royal Museum for Central Africa, Leuvensesteenweg 13, B-3080 Tervuren, Belgium

Abstract

We describe the isolation and characterization of 11 polymorphic microsatellite loci from the recently discovered fruit fly pest, *Bactrocera invadens*. The polymorphism of these loci was tested in individual flies from two natural populations (Sri Lanka and Democratic Republic of Congo). Allele number per locus ranged from three to 15 and eight loci displayed a polymorphic information content greater than 0.5. These microsatellite loci provide useful markers for studies of population dynamics and invasion history of this pest species.

Keywords: *Bactrocera invadens*, microsatellites, polymorphism

Received 23 May 2008; Revision accepted 7 July 2008

Asian fruit fly pests from the genus *Bactrocera* are regarded as some of the most destructive insects of fruits and vegetables worldwide (White & Elson-Harris 1992). In March 2003, three fruit fly specimens of the genus *Bactrocera* were detected in Kenya (Lux *et al.* 2003) and the species was described as *Bactrocera invadens* (Drew *et al.* 2005) in recognition of its rapid invasion of the African continent. It is feared that the geographical expansion of this pest and infestation could be more severe if no control programmes and efficient quarantine systems are implemented (Ekesi *et al.* 2006).

Microsatellite markers have been successfully applied to different invasive fruit fly species to infer evolutionary aspects underlying their invasive processes and to identify the routes of their colonization (Bonizzoni *et al.* 2004; Aketarawong *et al.* 2007). Here we report the characterization of 11 microsatellite markers in *B. invadens* and provide evidence of their usefulness for the study of genetic structure and dynamics of this species.

Correspondence: A. R. Malacrida, Fax: +390382986294; E-mail: malacrid@unipv.it

Genomic DNA was extracted from *B. invadens* flies from an International Center of Insect Physiology and Ecology (ICIPE) mass-rearing strain established in 2003 from wild flies collected from rotten mangoes in Nairobi (Kenya). Following digestion of genomic DNA with *Rsa*I (Roche), fragments were ligated to two adapter oligonucleotides (Adaptor A: 5′-CTCTTGCTTACGCGTGGACTA-3′; Adaptor B: 5′-TAGTCCACGCGTAAGCAAGAGCAC-3′) and used as template in polymerase chain reaction (PCR) with 24 pmol of adaptor A as primer. PCR products were denatured and stringently hybridized to a biotinylated $(dCA)_{10}$ probe. The target fragments were captured using streptavidin magnetic particles (Promega). Captured fragments were amplified, cloned and sequenced. A total of 200 plasmid inserts were analysed on an ABI PRISM 310 Genetic Analyser (Applied Biosystems). Twenty-six clones contained microsatellite sequences. Primers were designed for 11 loci with suitable flanking sequence using Primer 3 (Rozen & Skaletsky 2000). Preliminary PCR screening performed on 18 females and 27 males from the ICIPE strain yielded single locus amplifications, and the loci were polymorphic in both sexes. This excludes any condition of sex linkage for these loci. Amplifications were performed on a PTC-100

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd

Table 1 Characteristics of 11 microsatellite loci from *Bactrocera invadens* and their polymorphism in 54 male flies, from two wild populations

Locus	Repeat motif	Primer (5′–3′)	GenBank Accession	T_a (°C)	Allele size range (bp)	Total N_a	Sri Lanka (n = 27)			Democratic Republic of Congo (n = 27)		
							N_a	H_O	H_E	N_a	H_O	H_E
Bi1	$(CA)_9TT(CA)_2$	HEX CTCTTGACACTGGCCTCGTT R: GTATGGCCGGAGACATCAGT	EU868612	58	129–159	12	12	0.85	0.88	6	0.67	0.75
Bi2	$(CA)_7$	FAM GCACTCACTCAACCTTAGACG R: GCACTTGAATTGTGCGAAGT	EU868613	59	140–149	10	10	0.54	0.83	5	0.72	0.75
Bi3	$(TG)_5CTG$	FAM CGCGAATTTCAAGCATTTTT R: GGTCTTAAGGCCAAGCACAA	EU868614	60	130–132	3	3	0	0.14	2	0	0.07
Bi4	$(CT)_6GCT$	HEX GCACTCGCATGCTTGTAGTC R: CCGGTTTTGTGCGAAAAG	EU868615	59	123–126	4	4	0.78	0.74	4	0.48	0.70
Bi5	$(CA)_5GA(CA)_2$	HEX GCCAGTCAGTGTCTCGTCAA R: AGCGAGTTTGTTTTCGGTGA	EU868616	60	112–133	15	15	0.81	0.91	4	0.41	0.68
Bi6	$(CA)_3CC(CA)_3$	FAM GCGACAAGTTCGACACAAAA R: TACTGATTGTGCCGTGTGCT	EU868617	60	90–101	8	6	0.20	0.75*	6	0.58	0.67
Bi7	$(CA)_5A(CA)_2$	FAM CTCGCTCTTCATTCAATCCA R: CGACACGTTAAGTGGCAAAA	EU868618	58	107–118	5	5	0.56	0.49	3	0.22	0.31
Bi8	$(AC)_5AT(AC)_2$	FAM ACAAGTGCAGCAAAGACACG R: ATCACATCATGAGGCGTTCA	EU868619	61	118–138	7	6	0.59	0.57	4	0.63	0.62
Bi9	$(TG)_2TA(TG)_6$	HEX GCGCTGCTCGTAAACATCTA R: GGGCAAACACTTGGATTCAC	EU868620	61	92–108	9	8	0.74	0.75	5	0.85	0.74
Bi10	$(TG)_3AT(TG)_8$	FAM ATCGAGCAGATCACTGAGCA R: CCGCGAGTAGCAAATCTTTC	EU868621	56	142–155	11	7	0.63	0.83	7	0.63	0.71
Bi11	$(TC)_3T(TC)_3$	HEX TGGGTTCACCGTCCTTTAAT R: GCCCATAGACATCCAGGGTA	EU868622	60	140–147	3	3	0.22	0.26	2	0.19	0.17

HEX and FAM: Fluorescent dye labelled at the 5′ end of the forward primer. T_a, annealing temperature; N_a, number of alleles per locus; n, number of males tested; H_O and H_E, observed and expected heterozygosities, respectively; *deviations from Hardy–Weinberg equilibrium at $P < 0.05$ after Bonferroni correction.

thermocycler (MJR Inc.). Reactions for PCR consisted of 25 ng of genomic DNA, 1× reaction buffer, 2.5 mM $MgCl_2$, 25 μM dNTP, 0.5 U *Taq* polymerase (Invitrogen) and 5 μM of each primer, one of which was 5′ labelled with a fluorescent dye, in a total volume of 15 μL. PCR conditions were 2 min at 95 °C, 29 cycles of 30 s at 94 °C, 90 s at 56–61 °C and 90 s at 72 °C and 5 min of elongation at 72 °C. After loading on the Genetic Analyser, PCR products were analysed by the GeneScan program (Applied Biosystems). An individual was declared null (non-amplifying allele) after at least two amplification failures.

The 11 primers were tested to assess their polymorphism in two wild populations: Sri Lanka (Ranbukpitiya, 07°02′54″N, 80°30′52″E) and Democratic Republic of Congo (DRC, Kisantu Botanical Garden, 05°7′S, 15°5′E). Because only males are captured with methyl eugenol attractant, all the analyses were performed on male individuals, 27 from each population. All loci were polymorphic across the two samples (Table 1). The number of alleles per locus ranged from three to 15 and eight loci had a polymorphic information content PIC > 0.5 (Kalinowski *et al.* 2007). Genetic variability within populations was analysed using GenePop, version 3.4 (Raymond & Rousset 1995). A comparable high level of polymorphism was detected for the majority of loci in the two populations analysed, which also displayed the presence of private alleles. None of the loci displayed linkage disequilibrium (Fisher's exact test, GenePop). One case of deviation from Hardy–Weinberg equilibrium was observed at locus Bi6 after Bonferroni correction (Rice 1989) with heterozygote deficiency in the Sri Lankan population. The use of Micro-Checker, version 2.2.3 (van Oosterhout *et al.* 2004) indicates that this may be due to the presence of null alleles. This study provides the first set of microsatellite markers available for *B. invadens*. The range in repeat sequence length and diversity of these markers will allow the investigation of population structure and invasion history of this species.

Acknowledgements

This work was carried out within the African Fruit Fly Programme of ICIPE (supported by BMZ) and funds from the FORD foundation. We would like to thank Drs M. Billah and Taila for supplying samples of *B. invadens*.

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd

References

Aketarawong N, Bonizzoni M, Thanaphum S *et al.* (2007) Inferences on the population structure and colonization process of the invasive oriental fruit fly, *Bactrocera dorsalis* (Hendel). *Molecular Ecology*, **16**, 3522–3532.

Bonizzoni M, Guglielmino CR, Smallridge CJ, Gomulski LM, Malacrida AR, Gasperi G (2004) On the origins of medfly invasion and expansion in Australia. *Molecular Ecology*, **13**, 3845–3855.

Drew RI, Tsuruta K, White IM (2005) A new species of pest fruit fly (Diptera: Tephritidae) from Sri Lanka and Africa. *African Entomology*, **13**, 149–154.

Ekesi S, Nderitu P, Rwomushana I (2006) Field infestation, life history and demographic parameters of the fruit fly *Bactrocera invadens* (Diptera: Tephritidae) in Africa. *Bulletin of Entomological Research*, **96**, 379–386.

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program Cervus accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*, **16**, 1099–1006.

Lux SA, Copeland RS, White IM, Manrakhan A, Billah MK (2003) A new invasive fruit fly species from the *Bactrocera dorsalis* (Hendel) group detected in East Africa. *Insect Science and its Application*, **23**, 355–361.

van Oosterhout CV, Hutchinson WF, Wills DP, Shipley P (2004) Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

Raymond M, Rousset F (1995) GenePop (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Rice WR (1989) Analysis tables of statistical tests. *Evolution*, **43**, 223–225.

Rozen S, Skaletsky HJ (2000) Primer 3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology*, pp. 365–386. Humana Press, Totowa, New Jersey.

White IM, Elson-Harris MM (1992) *Fruit Flies of Economic Significance: Their Identification and Bionomics*. C.A.B International and Australian Centre for International Agricultural Research, Wallingford, UK.

KENYATTA UNIVERSITY LIBRARY

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd