The taxonomy and systematics of the bee genus *Scrapter* (Colletidae: Colletinae: Scraptrini)

Qiniso Michael Mthethwa

Submitted in fulfillment of the requirements for the degree of Master of Science (Genetics) in the School of Life Sciences, University of KwaZulu-Natal.

February 2016

Supervisor: Dr. Sandi Willows-Munro Signature:

Co-supervisor: Dr. Connal Eardley Signature:

ABSTRACT

Scrapter is a southern African endemic pollen-collecting genus consisting of 43 described species with 13 species also occurring in neighbouring countries such as Lesotho, Mozambique, Namibia and Zimbabwe. Classification of *Scrapter* has been partially investigated using only morphological data set. In this study I reassess previous classification and describe a putative new species which has never been described before, *Scrapter leovalis* sp. nov. This study aimed at reassessing previous classification of *Scrapter* using 46 morphological characters and provide a new classification based molecular data; 28SrRNA, EF-1α and Opsin gene. Using phylogenetic methods such as parsimony analysis, maximum likelihood and bayesian analysis classification of *Scrapter* was possible. However, analyses of morphological characters only did not produce reliable phylogenies due to low branch supports (bootstrap and posterior support). Molecular data was analyzed individually and in supermatrix analyses which produced phylogenetic trees with high branch supports. To see how the morphological and molecular data complemented each other a simultaneous analysis of 28SrRNA EF-1α, Opsin and morphological characters was carried out. Simultaneous analysis produced phylogenetic trees resembling supermatrix trees obtained in the simultaneous analysis of the three genes alone. It was then concluded that morphological characters chosen for this study were not phylogenetically informative.

DECLARATION-PLAGIARISM

I, **Qiniso Michael Mthethwa**, declare that:

1. The research reported in this thesis is my original work, and where the work of others has been used as a source it is acknowledged accordingly.
2. The experimental work and procedures described in this dissertation were conducted by me under the supervision of Dr Sandi Willows-Munro and field sampling was conducted under the supervision of Dr Connal Eardley
3. All assistance towards the production of this work and all the references contained herein have been duly accredited.

Student Name: Qiniso Michael Mthethwa

Signature: _____ **Date: _____**

Acknowledgements

I would like to send many thanks to my supervisor Dr. Sandi Willows-Munro and co-supervisor Prof. Connal Eardley who have been with me all this way, guided me, corrected me all the way and supported me emotionally and mentally. I would also like to thank Connal Eardley for lending me the dried specimen collection of this genus, which was used in this study. I would also like to thank Mr. Burgert Muller who helped with several species included in this study. I would love to give many thanks for the JRS Biodiversity bursary, UKZN School bursary and NRF funding which funded this study. I cannot begin to express my gratitude towards the Conservation Biology laboratory (Dr Sandi Willows-Munro's lab) members of 2012-2013. Thanks to all the people who have made this study possible, the whole UKZN PMB Genetics departments. And lastly I would love to thank my family for support during hard times of this research and long distance travelling during sampling seasons.

Table of Contents

Chapter 1: Introduction

Angiosperms and bees

Angiosperms have an estimated 250,000-260,000 species (Soltis and Soltis, 2004). Bees are the most crucial pollinators of angiosperms (Danforth *et al.*, 2006a) and as such they provide a major ecological service and contribute significantly towards the agricultural sector of most countries (Buchmann and Nabhan, 1996; Johannsmeier, 2001; Losey and Vaughan, 2006). There are currently over 19,000 described bee species globally (Michener, 2007; Bartomeus *et al.*, 2011; Ascher and Pickering, 2012) and most are dependent on close associations with angiosperm host plants (Danforth *et al.*, 2006a). This close relationship is best illustrated by the morphological specializations many bee species have which are specifically adapted for collecting, carrying, storing and using pollen and other plant products (Thorp, 1979, 2000). Bees are dependent on plant hosts for adult and larval nutrition, pollen, nectar and floral oils (Neff and Simpson, 1981; Buchmann, 1987; Wcislo and Cane, 1996), floral waxes and perfumes serving as sexual attractants (Dressler, 1982) as well as resins which are used to build nests (Armbruster, 1984). Many bee species specialize on one or a few closely related host plants (Wcislo and Cane, 1996). The recent decline in pollinators (including bees) is a worldwide threat for the pollination of many plant communities (Kevan and Phillips, 2001; Ashman *et al.*, 2004).

The bee genus *Scrapter*

Scrapter (Lepeletier and Serville, 1828) is a southern African endemic bee genus consisting of morphologically diverse bee species (Davies *et al.*, 2005). Approximately 40% of the species in the genus are distributed within the winter rainfall region of the Succulent Karoo Biome in the Western and Northern Cape provinces of South Africa (Davies *et al.*, 2005). The Succulent Karoo is an arid region which is exceedingly rich in endemic flora and fauna (Cowling and Hilton-Taylor, 1994; Milton *et al.*, 1997; Van Wyk and Smith, 2001; Smit *et al.*, 2008).

The *Scrapter* genus has been taxonomically placed within Colletinae (Michener, 2007). This family has recently been sub-divided based on morphology into three tribes; Paracolletini, Colletini and Scraptrini (Melo and Gonçalves, 2005); with the genus *Scrapter* placed within the tribe Scraptrini (Michener, 2007).

Eardley (1996) provided the first comprehensive taxonomic review of *Scrapter*. In this study he recognized 31 species in the genus and described 11 new species. A more recent study by Davies *et al*. (2005) provided another taxonomic revision of *Scrapter*, and recognized eight new species. In the following year, a further three new species were added to the genus (Davies and Brothers, 2006). *Scrapter* currently includes 43 described species (Eardley and Urban, 2010).

Vegetation

The floral preferences of *Scrapter* species have not been widely studied but members of six plant families (Aizoceae, Asteraceae, Neuradeceae, Proteaceae, Oxalidaceae and Zygophyllaceae) are considered to be pollinated by *Scrapter* species (Gess and Gess, 1993; Davies, 2006). The type species for the group, *S. bicolor* Lepeletier Saint-Fargeau & Audinet-Serville, has been collected in large numbers exclusively on *Herrea* species (Aizoceae). This shows the species is oligolectic to *Herrea*. However, there have been cases where other *Scrapter* species has been collected on Aizoceae flowers but those are rare occurrences (Davies, 2006). The *Oxalis* genus belongs to the Oxalidaceae family and is native to southern Africa and South America (Leistner, 2000). *Oxalis* comprises about 700 species world-wide (Leistner, 2000). Approximately 270 species occur in southern Africa. Some species are widespread, but the majority of taxa are confined to the winter rainfall Fynbos Biome in the Western Cape of South Africa (Leistner, 2000). Another plant known to be pollinated by *Scrapter* species is *Cotula barbata* DC, which belongs to the Asteraceae family. Asteraceae is the most frequented family with 11 *Scrapter* species recorded feeding on various composite genera (Davies, 2006). *Cotula barbata* is a South African endemic plant and mostly confined to the Northern Cape and Western Cape (Raimondo *et al.*, 2009).

Scrapter chloris, Scrapter luridus, Scrapter whiteheadi and *Scrapter avius* have been observed to frequent *Grielum* (Neuradaceae). Proteceae family is rarely visited by bees in southern Africa but two *Scrapter* species; *Scrapter erubescens* and *Scrapter fulginatus* forage exclusively on *Leucadendron* and *Paranomus* species (Davies, 2006). Species of *Zygophyllum* belonging to Zygophyllaceae are perennials and sometimes herbaceous annuals found mainly in the deserts and semi-deserts of the world (White, 1983). In Africa they are found in the arid and semi-arid areas in the southern part of the Eastern Cape, Western Cape and Northern Cape Provinces of South Africa, Namibia, southwestern Botswana and as far north as southern Angola (Low,

1998). Given that there is such a strong association between pollinators and the plants they pollinate, the distribution and evolutionary history of *Scrapter* is closely linked with the flora they pollinate (Davies *et al.*, 2005).

Utility of molecular data

The taxonomic classification of *Scrapter* based on morphology alone can be problematic because of the overlap of plesiomorphic morphological characters within members of the family Colletidae (Michener, 2007). *Scrapter* can be morphologically aligned with several other bee genera; and species belonging to *Scrapter* have been incorrectly assigned to other tribes and even subfamilies. For example, *Scrapter* has been placed into Paracolletini, a tribe sister to Colletini (Michener, 1944). McGinley (1981) provided evidence suggesting *Scrapter* should be sister to Euryglossinae, an Australian subfamily. The evidence presented in the latter study was based on three larval characters (relation of hypostomal ridge to head capsule; orientation of maxillary palpus; structure of salivary lips). However, these findings by McGinley (1981) were not supported by geological evidence as Africa and Australia have not been in contact for about 130 Million years (Early Cretaceous) and this date is much older than the origin of bees as a group ~125 Ma (Engel, 2001), so the proposed hypothesis seems unlikely.

Based on morphology, Melo and Gonçalves (2005) and Michener (2007) placed *Scrapter* in the tribe Scraptrini. Michener (2007) noted some distinguishing characters (a foveate prementum, reduced galeal comb and female basitibial plate being margined by a series of broken carinae), which can be used to separate *Scrapter* from most other hairy colletids. These characters suggest a relationship with the Euryglossinae instead of the Colletinae. In some species of *Scrapter* and Euryglossinae, the continuous carina forming an ordinary basitibial plate shows *Scrapter* to be related to Euryglossinae instead of Colletinae (Michener, 2007). Employing only morphological characters to understand the relationship between *Scrapter* to other colletids may not be useful as some characters are plesiomorphic; inherited from ancestor and not phylogenetically useful for resolving relationship among ingroup taxa. For instance, the narrow grooves like facial foveae are not only observed in *Scrapter* but this character is also present in some Hylaeinae, Euryglossinae and other Colletinae species (Michener, 2007). Consequently, the possession of the fovea on the prementun sometimes resolves *Scrapter* as the basal branch of the Hylaeine

clade, and not a member of Colletine in the study of short tongue bee families by Alexander and Michener (1995).

The morphological diversity observed among *Scrapter* species has led to different taxonomic treatments of the group. Eardley (1996) divided *Scrapter* into eight species-groups, which were convenient in describing the species but had no taxonomic relevance, whereas Michener (2000) in a subsequent study suggested two main groups within this genus. Michener (2000) considered species which have the following characteristics to belong to a group: (a) broad facial foveae of the female; (b) basitibial plate of the female with marginal carinae (or at least lower one) tuberculate or lobed; (c) propodeal triangle finely roughened; (d) thoracic sculpturing not especially coarse; (e) claws of female cleft or simple; and (f) body commonly larger and robust, with pale metasomal hair bands. Members of the second group are characterized by: (a) narrow facial foveae of the female; (b) basitibial plate of the female with simple marginal carinae; (c) propodeal triangle with striate dorsal surface separated from rest of the propodeum by pitted lines; (d) thoracic sculpturing of extremely coarse punctures, midline and notatauli deeply impressed on anterior end of scutum; (e) claws of female simple; and (f) body is small, slender, without metasomal hair bands.

Considering the confusion presented by the morphological classification of *Scrapter* the inclusion of molecular markers may shed some more light on the phylogenetic relationships within the genus. Given the previous findings relating to the traditional taxonomical classification of the *Scrapter* this study aims to test these previous morphological classifications of *Scrapter* through combined analysis of molecular sequences and morphological data within a phylogenetic framework. In the second chapter of this thesis all morphological data currently available for *Scrapter* is combined and analyzed. Using the morphological data a new species is described. In the third chapter molecular data from three genes is collected and used to construct a phylogeny. Lastly, the two types of data sets will be analyzed simultaneously to build a supermatrix phylogeny.

Chapter 2: Morphological analysis of *Scrapter* and description of a new species

Introduction

Scrapter Lepeletier & Serville, 1828 is a South African endemic bee genus belonging to the family Colletidae, subfamily Colletinae, and tribe Scraptrini (Eardley and Urban, 2010; Michener, 2007). *Scrapter* is a pollen-collecting genus with 43 described species, with members primarily distributed in South Africa; 13 species also occur in neighbouring countries such as Namibia, Mozambique, Zimbabwe and Lesotho (Eardley and Urban, 2010). This genus can be easily differentiated morphologically from the other Colletinae genera by five key features: the body is elongate, there are two submarginal cells in the fore wing (Davies and Brothers, 2006; Davies *et al.*, 2005; Eardley, 1996; Michener, 2000), the clypeus and paraocular areas lack yellow maculation (Davies and Brothers, 2006); the maxillary palpi are six-segmented and differ in length (Davies and Brothers, 2006); and facial foveae are observed in both sexes of all taxa (Davies and Brothers, 2006).

There have been conflicting viewpoints concerning the classification of species within *Scrapter*. Using morphological characters alone, Eardley (1996) divided *Scrapter* into eight conglomerate groupings, namely the nitidus (*Scrapter nitidus* (Friese), *Scrapter ruficornis* (Cockerell) and *Scrapter opacus* (Friese)), basutorum (*Scrapter basutorum* (Cockerell), *Scrapter flavipes* (Friese) and *Scrapter pallidipennis* (Cockerell)), flavostictus (*Scrapter flavostictus* (Cockerell), *Scrapter albitarsis* (Friese), *Scrapter absonus* (Eardley), *Scrapter avius* (Eardley), *Scrapter calx* (Eardley), *Scrapter caesariatus* (Eardley) and *Scrapter aureiferus* (Cockerell)), bicolor (*Scrapter bicolor* (Lepeltier and Serville), *Scrapter whiteheadi* (Eardley), *Scrapter niger* (Lepeltier and Serville), *Scrapter leonis* (Eardley), *Scrapter thoracicus* (Friese), *Scrapter tomentum*, *Scrapter algoensis* (Friese), *Scrapter albifumus* (Eardley), *Scrapter chloris* (Eardley) and *Scrapter luridus* (Eardley)), striatus (*Scrapter striatus* (Smith) and *Scrapter capensis* (Friese)), armatipes (*Scrapter armatipes* (Friese) and *Scrapter amplitarsus* (Friese)) and erubescens (*Scrapter erubescens* (Friese), *Scrapter fuliginatus* (Eardley) and *Scrapter amplispinatus* (Eardley)) species groups. The heterodoxus species-group (*Scrapter heterodoxus* (Cockerell)) is monotypic, with specimens of this species characterized by quite unique morphological characters including

a unique shape and coloration of the vestiture. These groupings were based on few characters shared between species in each grouping and will for the whole dissertation be referred to as "Eardley's groupings" to facilitate communication.

In a subsequent study, Michener (2000) proposed that *Scrapter* can be divided into two major groups based on facial foveae, basitibial plate and pretarsal claws of the female, propodeal triangle sculpturing, thoracic sculpturing, and difference in body sizes. The first group consists of *S. erubescens*, *S. fuliginatus*, *S. chloris*, *S. amplispinatus*, *S. striatus*, *S. capensis*, *S. carysomus*, *S. catoxys*, *S. eremanthedon*, *S. tomentum*, *S. algoensis*, *S. basutorum*, *S. pallidipennis*, *S. flavostictus*, *S. albitarsis*, *S. absonus*, *S. calx*, *S. caesariatus*, *S. aureiferus*, *S. heterodoxus*, *S. bicolor*, *S. whiteheadi*, *S. leonis*, *S. thoracicus*, *S. niger*, *S. amplitarsus* and *S. armatipes*. The second group consists of *S. pyretus*, *S. sittybon*, *S. viciniger*, *S. acanthophorus*, *S. chrysomastes*, *S. glarea*, *S. nitidus*, *S. ruficornis* and *S. opacus*. However, in a few species these morphological characters are not consistent, i.e., some species show characteristics of both groups (e.g. *S. flavipes*, *S. avius*, *S. rufescens* (Friese), *S. luridus* and *S. albifumus*), and assignment to these two groupings is often unclear (Michener, 2000). The other two *Scrapter* species making a total of 43 species were not included in Michener's (2000) groupings.

Davies (2006) conducted a cladistic analysis using 25 morphological characters and 42 species to evaluate the conflicting viewpoints. In this study he observed three highly supported clades: nitidus (*S. ruficornis*, *S. opacus*, *S. nitidus*, *S. pallidipennis* and *S. flavipes*), erubescens (*S. erubescens*, *S. fuliginatus* and *S. amplispinatus*) and chloris (*S. chloris*, *S. luridus* and *S. whiteheadi*) clades. The rest of the relationships within the genus were not well supported. The results only provided partial support for the divisions proposed by Eardley (1996) and Michener (2000).

In this chapter I aim to extend the morphological character sampling used by previous authors Eardley (1996), Michener (2000) and Davies (2006), and reanalyze the morphological data to provide clarification on the classification of the *Scrapter* species.

Materials and methods

In total 46 morphological characters were analysed for 38 species (35 described *Scrapter* species, one putative new species and two outgroups, *Leioproctus irroratus* and *Leioproctus plumosus*

representing Paracolletinae). Due to a limited sampling period and not having access to all the dried material, seven species were missing from the analysis (*Scrapter acanthophorus*, *Scrapter carysomus*, *Scrapter fulginatus*, *Scrapter glarea*, *Scrapter rufescens*, *Scrapter sittybon* and *Scrapter viciniger*). Specimen information was obtained through examination of fresh collected material, archival material from the South African National Collection of Insects (SANC) reference collection or sourced from the literature (details below).

Sample collection

Qiniso Mthethwa with the help of Dr. Connal Eardley collected 87 *Scrapter* specimens representing 17 species (*S. algoensis*, *S. amplispinatus*, *S. armatipes*, *S. bicolor*, *S. caesariatus*, *S. catoxys*, *S. chloris*, *S. eremanthedon*, *S. erubescens*, *S. flavipes*, *S. luridus*, *S. niger*, *S. nitidus*, *S. opacus*, *S. ruficornis*, *S. tomentum* and *S. whiteheadi*) in the Northern Cape, Western Cape and KwaZulu-Natal provinces (Figure 1). Localities in the Northern Cape and Western Cape were visited in September – October 2012, while localities in KwaZulu-Natal region in January 2013. In addition to field collected specimens, 31 museum specimens were examined (Table 1). The collection localities for specimens which have associated GPS co-ordinates are also shown on Figure 1. Seven museum records did not have GPS coordinates and are therefore not shown in Figure 1. Morphological characters for *S. pruinosus* and *S. pyretus* were coded from literature.

Specimens caught in the field were collected using sweep nets and killed with cyanide. Representative specimens for each of the identified species listed above were stored in 99.99% ethanol in the -20^{o}C freezer for DNA extraction later, whereas the rest of the collected specimens were pinned for museum storage and deposited in the South African National Collection of Insects (SANC). Species identification was done in collaboration with Dr Connal Eardley using the reference collection housed at the SANC in Pretoria, South Africa. The putative new species *Scrapter leovalis* sp. nov was borrowed from the KwaZulu-Natal Museum in Pietermaritzburg, South Africa. This putative new species has not been formally described in previous studies. Specimen belonging to 25 *Scrapter* species were borrowed from the SANC collection (Table 1). Two specimens of each species (male and female, where possible) were included in the analyses. The specimens collected in the present study (noted by * in Table 1) have not received the database numbers, thus not available (N/A) as shown in Table 1.

Table 1. The names and collection localities of the specimens used for coding morphological characters in this study. Accession numbers (Accession No.) refer to the identification number for each specimen placed in the SANC. F = female, M = male, Nam = Namibia, S.A = South Africa, NCP = Northern Cape Province, O.F.S. = Orange Free State, WCP = Western Cape Province, E = East, S = South, NE= North East, Nat = Nature, Nati= National, Res = Reserve, SANC = South African National Collection of Insects, N/A = information not available. Asterisk (*) indicate specimen collected in the present study. Need to add in two extra columns listing collection date and collector following examiners comments.

Sp. Name & Sex	Locality and Date	GPS Coordinates	Museum	Accession No.	Collector
Outgroups					
L. irroratus, M+F	Australia. NSW, Hilltop. 2.xii.1999	N/A	N/A	Leir705	N/A
L. plumosus	Australia. Victoria, Torquay. 19.xi.1999	N/A	N/A	Lepl706	N/A
Ingroups					
S. absonus, M+F	S.A. Langjan Nat. Res.10.iii.1990	22.52S, 29.14E	SANC Pretoria	HYMA04059	C.D. Eardley
S. albifumus, M+F	S.A. NCP, Richtersveld Nati. Park.12.ix.2001	28.18S, 16.58E	SANC Pretoria	HYMA08484	C.D. Eardley
S. albitarsis, M	S.A. Namaqualand.5.vi.1985	N/A	SANC Pretoria	HYMA08495	M. Stuck
S. albitarsis, F	S.A. WCP, 4m S Elands Bay.18.ix.2005	32.25S, 18.20E	SANC Pretoria	HYMA08507	C.D. Eardley
*S. algoensis**, M+F	S.A. NCP, Springbok.11.ix.2012	29.34S, 18.01E	SANC Pretoria	N/A	Q.M. Mthethwa
*S. amplispinatus**, M	S.A. NCP, Springbok.12.ix.2012	29.34S, 18.01E	SANC Pretoria	N/A	Q.M. Mthethwa
S. amplispinatus, F	Not labeled	N/A	SANC Pretoria	N/A	
S. amplitarsus, M	S.A. NCP, Springbok, Goegap Nat. Res.13.viii.1993	N/A	SANC Pretoria	HYMA04077	V.B. Whitehead
S. amplitarsus, F	S.A. Namaqualand.04.ix.1985	N/A	SANC Pretoria	HYMA04080	M. Stuck
*S. armatipes**, M+F	S.A. NCP, Springbok.11.ix.2012	29.34S, 18.01E	SANC Pretoria	N/A	N/A
S. aureiferus, M	S.A. WCP, Sauer, 26 km Piketberg.21.x.2008	32.50S, 18.33E	SANC Pretoria	N/A	C.D. Eardley
S. aureiferus, F	S.A. NCP, Wallekraal.x.1950	N/A	SANC Pretoria	HYMA04069	N/A
S. avius, M	S.A, NCP, Richtersveld Nati. Park.23.viii.2006	28.03S, 17.03E	SANC Pretoria	HYMA08496	C.D. Eardley
S. avius, F	Nam. Boom river Canyon, 4km of Orange river.25.x.1996	28.00S, 17.03E	SANC Pretoria	HYMA08501	N/A
*S. bicolor**, M+F	S.A. NCP, Springbok.11.ix.2012	29.34S, 18.01E	SANC Pretoria	N/A	Q.M. Mthethwa
S. basutorum, M+F	S.A. O.F.S. Adullam Farm near Clarens	28.34S, 28.28E	SANC Pretoria	HYMA08515	N/A
*S. caesariatus**, M+F	S.A. NCP, Garies.14.ix.2012	30.13S, 18.07E	SANC Pretoria	N/A	Q.M. Mthethwa
S. calx, M	S.A. Cape Agulhas.27.ix.2001	34.50S, 20.01E	SANC Pretoria	HYMA08696	C.D. Eardley
S. calx, F	S.A. WCP, Vanrhysdorp.06.ix.2007	28.23S, 18.40E	SANC Pretoria	HYMA08685	C.D. Eardley
*S. catoxys** M+F	S.A. NCP, Springbok.12.ix.2012	29.34S, 18.01 E	SANC Pretoria	N/A	Q.M. Mthethwa
S. capensis, M+F	S.A. Namaqualand.05.vi.1985	N/A	SANC Pretoria	HYMA04098	M. Stuck
*S. chloris**, M+F	S.A. NCP, Garies.14.ix.2012	30.12S, 18.07 E	SANC Pretoria	N/A	Q.M. Mthethwa
S. chrysomastes, M	S.A, NCP, Richtersveld Nati. Park.23.viii.2006	28.03S, 17.03E	SANC Pretoria	HYMA08530	C.D. Eardley

Table 2. (Continued)

Sp. Name & Sex	Locality	GPS Coordinates	Museum	Database No.	
*S. eremanthedon** M+F	S.A, NCP, North Eksteenfontein.13.ix.2012	28.49S, 23.10 E	SANC Pretoria	N/A	Q.M. Mthethwa
*S. erubescens**, M+F	S.A. NCP, Garies.14.ix.2012	30.11S, 18.10E	SANC Pretoria	N/A	Q.M. Mthethwa
*S. flavipes**, M+F	S.A. NCP, Springbok.11.ix.2012	29.34S, 18.01E	SANC Pretoria	N/A	Q.M. Mthethwa
S. flavostictus, M+F	S.A. Natal Cathedral Peak Area.28.iii.1985	28.59S, 29.14E	SANC Pretoria	HYMA04111	C.D. Eardley
S. heterodoxus, M	S.A. WCP, Hermanus.29.ix.2001	34.14S, 18.26E	SANC Pretoria	HYMA08537	C.D. Eardley
S. heterodoxus, F	S.A. WCP, Citrusdal.07.ix.2001	32.36S, 18.53E	SANC Pretoria	HYMA08539	C.D. Eardley
S. leonis, F	S.A. Capland Wllowmore.15.viii.1921	N/A	SANC Pretoria	HYMA04126	H. Brauns
S. leovalis sp. nov, M+F	S.A. WCP, Nieuwoudtville, Farm Glen Lyon, Renosterveld.11.ix.2003	31.24S, 19.08E	KwaZulu-Natal Museum (PMB)	N/A	M. Kuhlman
*S. luridus**, M+F	S.A. NCP, Garies.14.ix.2012	30.27S, 17.41E	SANC Pretoria	N/A	Q.M. Mthethwa
*S. niger**, M+F	S.A. NCP, Garies.14.ix.2012	30.27S, 17.41E	SANC Pretoria	N/A	Q.M. Mthethwa
*S. nitidus**, M+F	S.A. NCP, Garies.14.ix.2012	30.27S, 17.41E	SANC Pretoria	N/A	Q.M. Mthethwa
*S. opacus**, M+F	S.A. NCP, Springbok .11.ix.2012	29.34S, 18.01E	SANC Pretoria	N/A	Q.M. Mthethwa
S. pallidipennis, M	S.A. NE kleinmond.29.ix.2001	34.20S, 19.05E	SANC Pretoria	HYMA08595	C.D. Eardley
S. pallidipennis, F	S.A. Dassiefontein Farm	30.09S, 17.59E	SANC Pretoria	N/A	
S. pruinosus, M+F	Nam. Lüderitz.7.xii.1994	26.35S, 15.07E	SANC Pretoria	N/A	M. Kuhlmann
S. pyretus, M+F	Nam. Swakop River.19.iii.1997	22.41S, 14.35E	SANC Pretoria	N/A	S.K. Gess
*S. ruficornis**, M+F	S.A. NCP, Springbok .11.ix.2012	29.34S, 18.01E	SANC Pretoria	N/A	Q.M. Mthethwa
S. striatus, M	Not labeled	N/A	SANC Pretoria	N/A	
S. striatus, F	S.A. WCP, Villiesdorp.20.ix.2001	33.58S, 19.16E	SANC Pretoria	HYMA08604	C.D. Eardley
S. thoracicus, M	S.A. Namaqualand, Hester Nalan.26.vi.1985	N/A	SANC Pretoria	HYMA04181	M. Stuck
*S. tomentum**, M+F	S.A. NCP, Garies.13.ix.2012	30.27S, 17.41E	SANC Pretoria	N/A	Q.M. Mthethwa
*S. whiteheadi**, F	S.A. NCP, North Eksteenfontein.13.ix.2012	28.54S, 16.44E 16.44S, 28.54E	SANC Pretoria	N/A	Q.M. Mthethwa

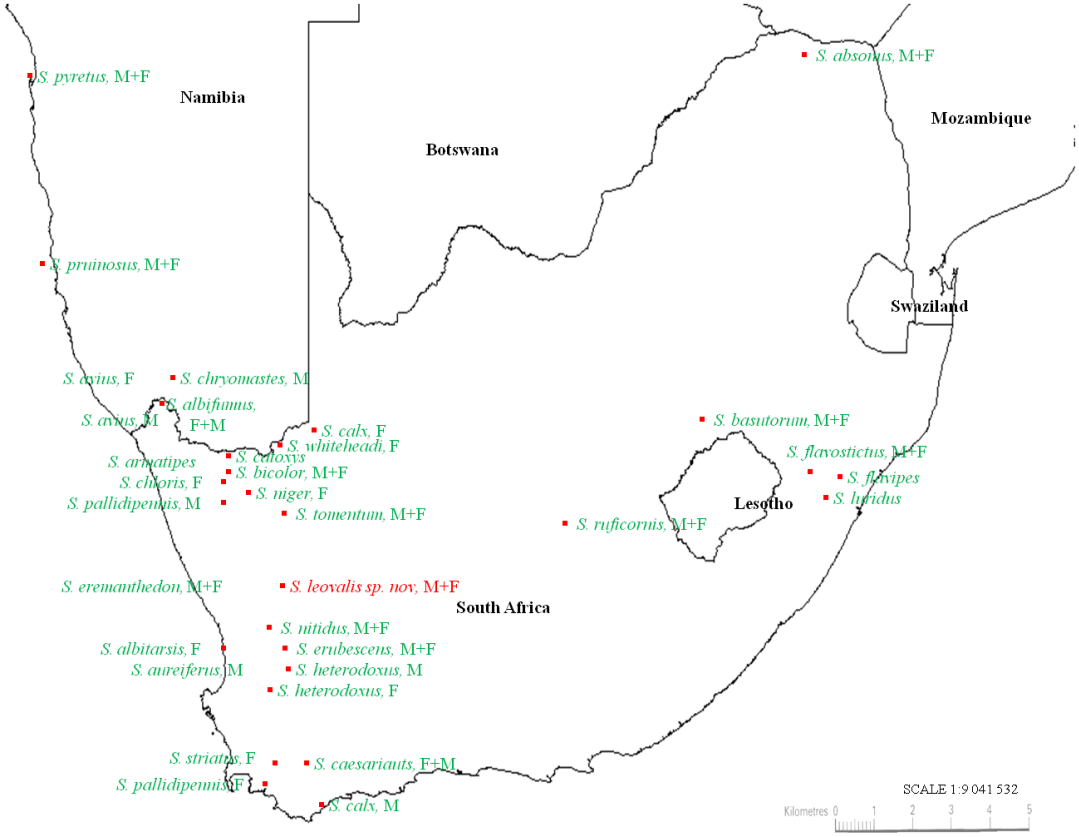

Figure 1. Geographic distribution of *Scrapter* specimens examined for morphological data analysis. The putative new species is written in red on the map. GPS coordinates for *S. pruinosus* and *S. pyretus* were taken from the literature.

Morphological data

There was a total of 38 species available for morphological character coding and analyses. Four species (*S. chrysomastes, S. leonis, S. thoracicus,* and *S. whiteheadi*) had >50% missing data. To test the impact of missing data on the analyses, two data sets were constructed; the first data set (Appendix 2; Table 3) contained data from all 38 species including the species with missing data. The second data set (Appendix 2; Table 4) excluded the four species with missing data and included only 34 species.

The study included a total of 46 characters (listed and described in Appendix 1), 40 which have been used in previous studies (Eardley, 1996; Michener, 1944, 2007; Engel, 2001; Davies and Brothers, 2006) and 6 new characters were added. New characters included; mandible colour, antennal scape colour, fovea on the second tergum (T2), fore tibia colour in male specimens, gonocoxa structure and basitibial plate in female specimens. The chosen morphological characters have been used in distinguishing certain groupings within the genus (Eardley, 1995; Michener, 2000), and could be phylogenetic informative. The morphological characters

showing high intraspecific variation (e.g. body length or color variation within single species) were avoided for this study. Diagrams of morphological characters used in the present study are shown in Appendix 1 with character states description. These were adopted for the present study from characters used in previous studies (Eardley, 1995; Davies *et al.,* 2005; Davies and Brothers, 2006).

Where specimens were in good condition, two specimens of each sex were examined. Eighteen characters were coded on both sexes, fourteen characters were coded only in male specimens and sixteen characters were coded only in females. In some cases reference material was not available and morphological information was sourced from the literature for two species (*S. pruinosus* and *S. pyretus*; Davies, 2006). Where specimens were damaged and characters could not be coded, those characters were treated as missing data (?) in the analysis.

Phylogenetic analyses

The two data sets were analyzed using parsimony, maximum likelihood (ML) and Bayesian optimality criteria. The parsimony analysis was carried out in TNT program (Goloboff *et al.*, 2000). The analysis was conducted out using a traditional search with 1000 replicates and 10 starting trees per replicate. Tree bisection algorithm (TBR) swapping algorithm was used, saving 10 trees per replication. Equally parsimonious trees were used to create consensus trees by selecting the "estimate consensus" option. ML analysis was carried out in GARLI v2.0 (Zwickl, 2006) using the Markov variable (Mkv) model as implemented by Lewis (2001). Model-fit was improved by dividing data into subsets based on the number of states e.g. two-state characters were analyzed in the same partition. Two runs for each data set (38 species and 34 species) were conducted. The first run was conducted with five replicate searches without bootstrap replicates to find the most likely tree. The second run was conducted consisting of 100 bootstrap replicates.

Bayesian inference (BI) was performed in MrBayes v3.2 (Huelsenbeck and Ronquist, 2003). Morphological data sets can be tricky to analyze in the MrBayes program as model selection is difficult. Four evolutionary models were tested (Table 2) and compared using Bayes factors in Tracer v1.5 (Rambaut and Drummond, 2009). The Bayes factor (BF) is a statistical Bayesian alternative to frequentist hypothesis testing that is often used for multiple model comparison (Rouder and Morey, 2012). The model implemented in MrBayes allows for a gamma shaped rate variation and variable coding. The models tested were based on this model, changing only two parameters; rates and coding. Interpretation of model support by Bayes factor was based on the guide lines described in Jeffreys (1961). Model 1 was selected as the best fit model using Bayes factors for both datasets (see Table 2) and was used in all BI runs.

Four separate runs were performed for each data set to verify that the resulting tree topologies were not biased due to the different starting points (Danforth *et al.*, 2006b) and to check the consistency of the obtained

topologies. Each analysis consisted of running four simultaneous chains for 2 million generations, sampling every 1000th generation. After each Bayesian run, convergence was assessed using Tracer v1.5 (Rambaut and Drummond, 2009) by observing the trace plots, density curves and effective sample size (ESS: Kass *et al.*, 1998). The first 2500 trees were removed as burnin from the tree files. Consensus trees were generated using the Consense module in the PHYLIP package (Felsenstein, 2004). Phylogenetic trees were visualized in FigTree v1.3.1 (Rambaut, 2009). All phylogenetic trees were rooted using *L. irroratus* and *L. plumosus*.

Table 3. Models compared using Bayes factors.

Model	**Parameters**		**34 taxa**	**38 taxa**
	Among site rate heterogeneity	**Coding**	**Bayes factor (log10)**	**Bayes factor (log10)**
Model 1	Gamma	Variable	259.165	257.154
Model 2	Gamma	All	0.835	0.810
Model 3	Equal	Variable	1.198	1.080
Model 4	Equal	All	0.001	0.001

Results

Generally, trees recovered by all three optimality criteria (parsimony, ML and BI) were poorly supported (Figure 2, 3 and 4). In parsimony and ML trees bootstrap support values less than 50% and Bayesian posterior probability values less than 0.80 are not shown on the trees. The ML and BI recovered similar topologies, while the topology recovered by the parsimony analysis was unresolved (Figure 2). The analysis of 38 and 34 taxa did not change the phylogenies recovered by any of three analyses.

Parsimony analysis

Parsimony analysis of the 38 taxa data set recovered 10 most parsimonious trees (tree length 233, CI = 0.27 and RI = 0.49). The consensus tree was not resolved with no branches supported by >70%.. Parsimony analysis of the truncated data set (34 taxa) resulted in four most parsimonious trees (tree length 215, CI = 0.30 and RI = 0.49) and the consensus tree was again not resolved (Figure 2b). Excluding taxa with missing data in the analysis did not have a significant effect on the parsimony analysis as the there were no resolved clades in the phylogeny Figure 2b.

Maximum likelihood analysis

Model-based analysis also resulted in a poorly resolved tree (Figure 3). Some associations were, however, supported. *S. opacus* was placed at the base of the tree with 65% and 95% (Figures 3a and b), followed by *S. caesariatus* and *S. eremanthedon* branch (52% in Figure 3a and 50% in Figure 3b) in the *S. nitidus* clade

consisting of *S. nitidus*, *S. caesariatus* and *S. eremanthedon*. The *S. whiteheadi* and *S. leonis* branch was well supported with 82% bootstrap support (Figure 3a). However, these taxa had more than 50% missing data. Analysis of the truncated data set (Figure 3b) did not produce a better or a unique tree, however, the placement of *S. opacus* at the base of the tree received rather improved 95% bootstrap support and the association of *S. pruinosus* and *S. pyretus* also showed an increased 86% bootstrap support. The placement of *S. leovalis* sp. nov is consistent in Figures 3a and b. The clade of *S. eremanthedon* and *S. caesariatus* remained poorly supported (50% and 52%) in both 38 and 34 taxa analyses unlike the *S. pyretus* + *S. pruinosus* clade showed improvement in Figure 3b with 86% bootstrap support.

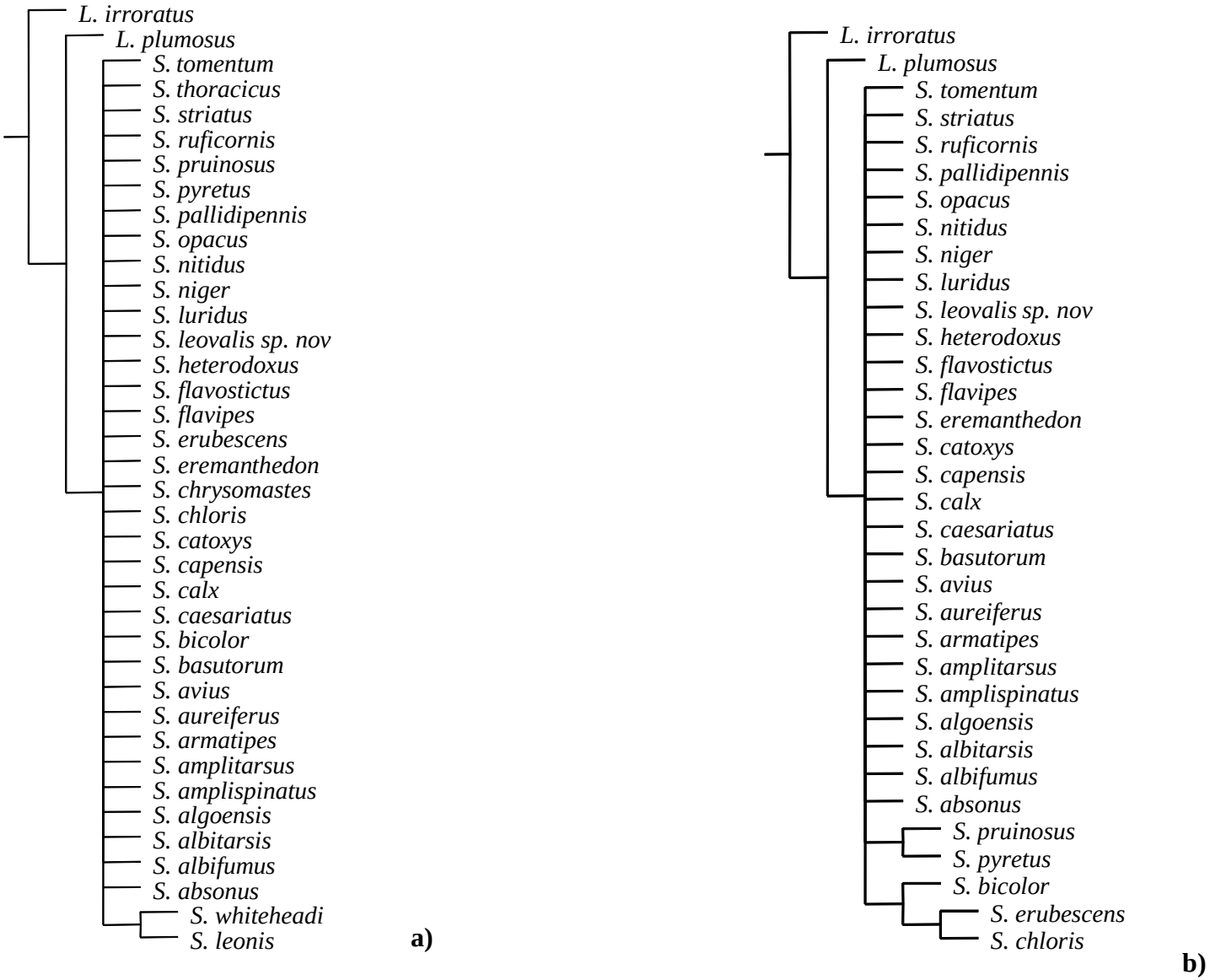

Figure 2. (a) Strict consensus phylogeny showing results of parsimony analysis of the 38 taxa. (b) Strict consensus phylogeny showing results of parsimony analysis of the 34 taxa. No bootstrap values above 50% were recovered.

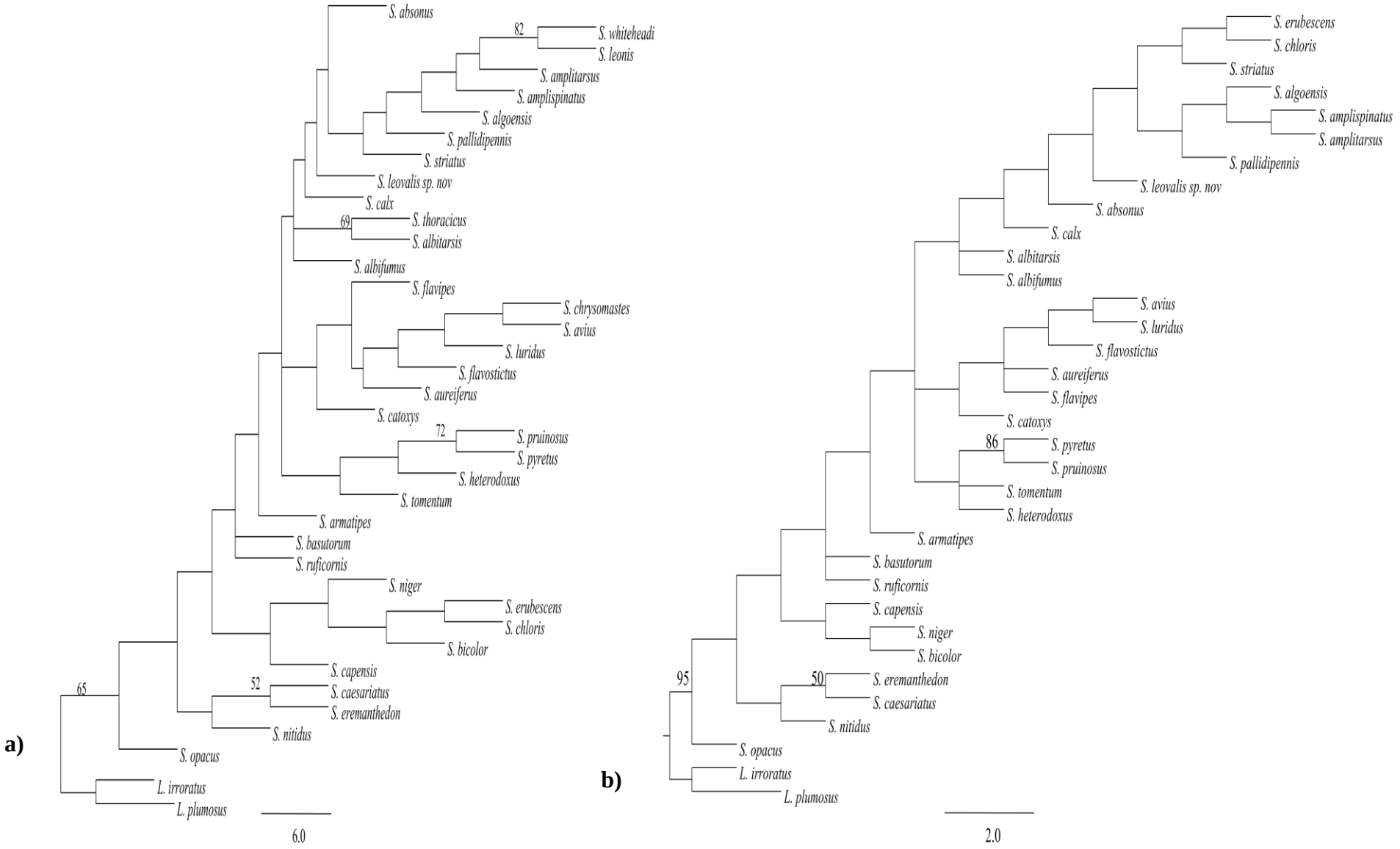

Figure 3. Maximum likelihood analyses of the (a) 38 taxa matrix and (b) 34 taxa matrix. Bootstrap support values are shown above the branches. However, bootstrap support values below (50%) are not shown.

Bayesian analysis

Phylogenetic trees recovered from the Bayesian analysis did not receive high posterior probability support values (Figure 4, values less than 0.80 posterior probability are not shown on the trees). Only the *S. pruinosus* and *S. pyretus* branch received posterior probability support of 0.92 and 0.96 in Figure 4a and b respectively. The association of *S. pyretus* and *S. pruinosus* was consistent in ML and BI analyses with acceptable bootstrap support in ML (Figure 3) and acceptable posterior probability support in BI (Figure 4). The position of the undescribed species *S. leovalis* sp. nov was consistent in Figures 4a and 4b as it was in Figures 3a and b. The relationship of several species was consistent regardless of the poor posterior probability support (not shown), i.e. the *S. bicolor* clade (*S. bicolor*, *S. calx* and *S. erubescens*) were constantly grouped together albeit with no support. Other branches that were consistent in Figures 4a and b include the *S. eremanthedon* and *S. caesariatus*; *S. algoensis* clade (*S. algoensis*, *S. striatus* and *S. pallidipennis*). Generally, Figure 4a and 4b displayed the same taxonomic classification of the species within the *Scrapter* genus.

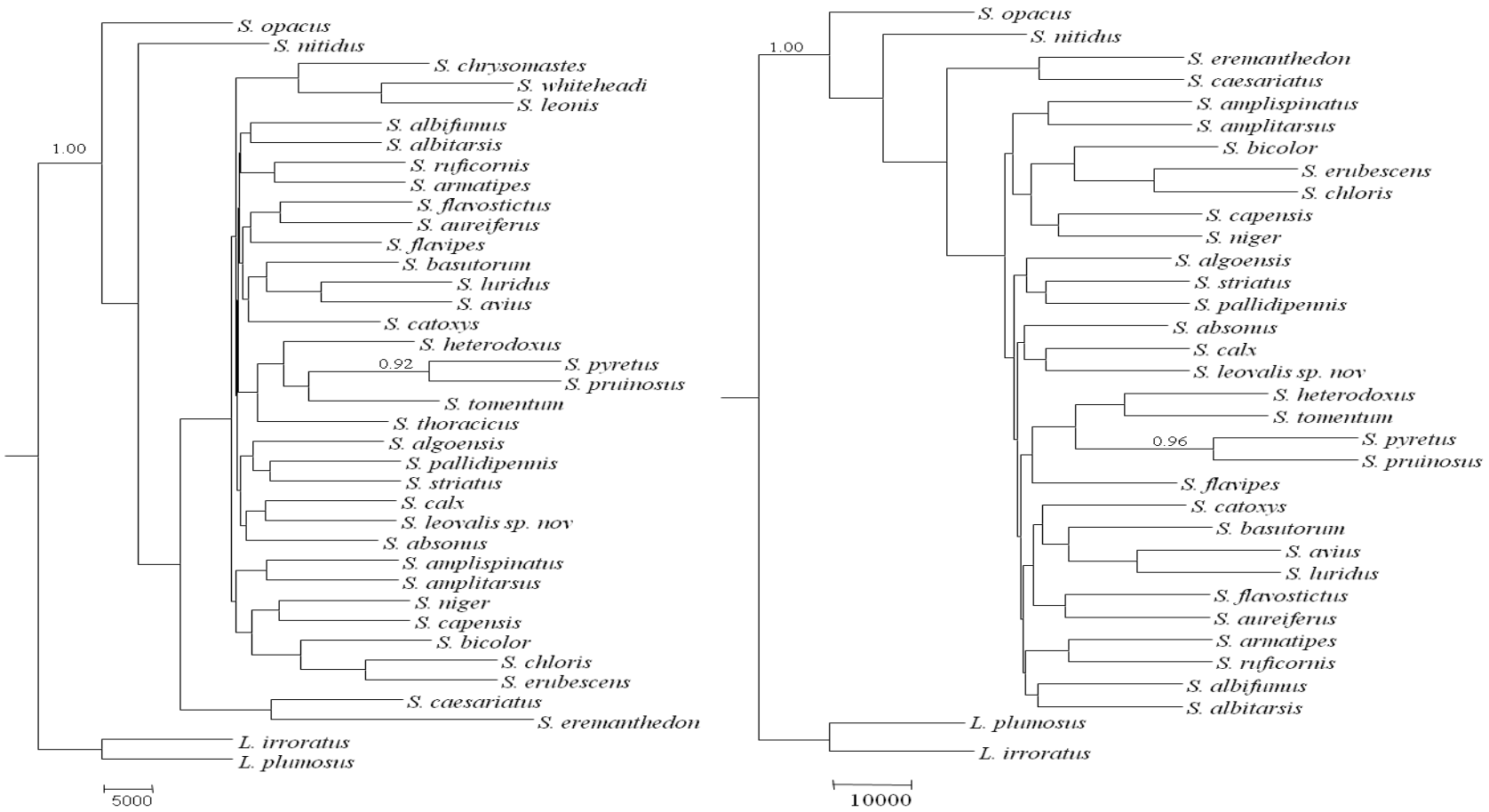

Figure 4. Bayesian consensus phylogeny of *Scrapter* species and outgroup analyzed in this study, a) 38 taxa and b) 34 taxa data sets. Posterior probabilities are indicated for each node when > 0.80.

Discussion

Any heritable characters can be used to infer phylogeny. Traditionally morphological characters have played a large role in the taxonomy of the southern African bee genus *Scrapter* (Eardley, 1996; Davies, 2006; Michener, 2007). In this chapter I present a comprehensive reanalysis of 46 morphological characters using a phylogenetic approach. In total 46 morphological characters were scored for 38 taxa. In general the phylogenies presented were poorly resolved with most associations not well supported. The addition of missing data had no significant effect on the phylogenetic reconstruction although in some cases branch support declined see Figure 3a, b and Figure 4a and b. All three methods of phylogenetic reconstruction (parsimony, maximum likelihood and Bayesian inference) resulted in largely unresolved phylogenies. The model-based methods did, however, provide some clarification.

A lineage including *S. pyretus* and *S. pruinosus* was well supported by moderate ML bootstrap support (72% in Fig 3a and 86% in Fig 3b) and posterior probability support (0.92 in Fig 4a and 0.96 in Fig 4b). This association was also noted by Davies (2006) who described several synapomorphic characters for these taxa, for example, antennae which are brownish dorsally and yellow ventrally, mandibles which are distally black and basally yellow and a clypeus which is mostly densely punctate with smooth and shiny interspaces and without a medio-longitudinal sulcus. In addition, the propodeum is angulate and the pygidial plate is absent in the males, while the female specimens have different vestiture appearances but their integument colour is similar, being black. The *S. bicolor*, *S. erubescens* and *S. chloris* clade was also consistent in ML and BI analytical methods. Morphologically these taxa also supported by the presence of fovea in the second tergum, shiny facial fovea and a largely reddish metasoma.

The effects of missing data have been debated in previous investigations (Novacek, 1992; Wilkinson, 1995; Wiens, 2003) and in this study the inclusion of missing data had no impact on the resolution of the phylogenies, particularly the parsimony and maximum likelihood phylogenies. Missing morphological data was considered misleading when reconstructing phylogenetic relationships (Anderson, 2001; Donoghue *et al.*, 1989; Huelsenbeck, 1991; Wiens, 2003) and in combined data sets (e.g. genes and morphology) that do not include identical taxa (Wiens and Reeder, 1995; Sanderson *et al.*, 1998). It can lead to an increased number of shorter trees and decreased resolution (Wiens and Reeder, 1995). Huelsenbeck (1991) reported that

highly incomplete taxa sampling can decrease phylogenetic accuracy because missing data increase the percentage of equivocally resolved ancestral characters, which leads to decreased resolution and thus to decreased phylogenetic accuracy. However, subsequent studies by Wiens (2003) and Wiens and Morrill (2011) provided evidence that data sets with up to 90% missing data for a few particular taxa can be correctly and efficiently used to provide substantial support for analyses. Although missing data is certainly a factor in this study, I do not believe that the presence of missing data biases phylogenetic estimation as removing the taxa with missing data had very little impact on the resolution in the trees.

The poor support observed in the phylogenies constructed using morphological characters alone is probably due to a lack of data, i.e., a limited number of morphological characters coded, low characters/taxa ratios and the missing data in the matrix. It cannot be concluded that morphological data alone can or cannot fully resolve the evolutionary history of this group, rather, from the results of the three analyses it can be concluded that the morphological characters sampled in this study were not sufficient for phylogenetic analyses. In the next chapter molecular data will be used to complement these results.

The morphological data does, however, provide evidence for a new species. The new species, *Scrapter leovalis* sp. nov, is described below. Additional analyses will need to be conducted to clarify the position of the taxon within *Scrapter*.

Taxonomy

Description of *Scrapter leovalis* sp. nov

Specimens belonging to the putative new species used in this paper are the property of the KwaZulu-Natal Museum, South Africa. Samples were examined and measured under a stereo microscope in UKZN Electron Microscope department, South Africa. The terminology follows that of Michener (2007). Bilaterally symmetrical structures are described in the singular. Abbreviations used: F- flagellomere, T- metasomal tergum and S- metasomal sternum. The term

carinulate in describing surface sculpture refers to largely parallel, raised, longitudinal ridges that are fairly densely packed together.

Species Diagnosis

Both female and male specimens of the species can be recognized by white sparse vestiture, ventrally yellow antennal flagellum, metasoma brownish, no apical hair bands on tergum 1-4 , stigma wide, propodeum strongly angulate, mesoscutum with sparse punctation, head and mesosoma shiny black, propodeal triangle forming an approximately equilateral triangle. Antennal flagellum is brownish to yellow ventrally. There is sparse punctation on the mesoscutum of both female and male specimens. And the species has no clypeal mediolongitudinal sulcus.

Description

Female:

3 Paratypes: South Africa, Nieuwoudtville Farm Glen Lyon, Renosterveld S31.24.03 E19.08.34 700m M. Kuhlmann leg. 11. ix. 2003

Measurements (*n*= 3): head length 1.90-2.30 mm, head width 1.15-1.72 mm, length of clypeus 0.4-0.6 mm, lower interocular distance 1.00-1.20 mm, upper interocular distance 1.15-1.28 mm, length of eye 1.00- 1.20 mm, interantennal distance 0.40-0.50 mm, antennocellar distance 0.38-0.41 mm, antennocular distance 0.30-0.40 mm, length or malar area 0.10-0.12 mm, mesoscutum length 1.00-1.20 mm, mesosoma length 1.80-2.00 mm, length of stigma 0.80-1.00 mm, width of stigma 0.16-0.20 mm, length of marginal cell beyond stigma 0.75-0.80 mm, length of marginal cell 1.00-1.30 mm, length of free part of marginal cell 0.72-0.80 mm.

Head: clypeus slightly convex, integument between punctures glabrous, sparse punctures about two to three puncture diameter apart, no clypeal mediolongitudinal sulcus. Mandible black with reddish apex, ventral side with sparse vestiture. Supraclypeal area glabrous between punctures, greatly convex, sparse punctation two-puncture diameter apart. Scape black, sparse white vestiture. Vertex seen from front, weakly convex to almost flat, almost the same level with summit of eyes. Frontal line carinate. Compound eyes parallel ventrally. Gena area finely sculptured, shiny, white sparse vestiture. Mouth parts not exposed. Frons sparse punctation about two puncture diameter apart. Facial foveae sulcus-like, shiny. F1-F11 yellow ventrally, dorsally

black. **Mesosoma:** pronotum shiny. Mesocutum and scutellum with finely reticulate sculpture. Mesoscutum with sparse punctation approximately two to three puncture diameter apart, sparse white vestiture. Propodeum strongly angulate, basal area roughly sculptured to carinate, naked. Pronotal lobe brownish to black. **Metasoma:** T1-T3 brownish, T4 brownish anteriorly, black posteriorly. T5-T6 black. T1-T3 graduli discs shiny, reticulate sculpture. **Legs:** fore tibia mostly brown to black with yellow spot dorsally. Mid and hind tibia brown to black. All tarsi brown to black. Metabasitibial plate not entire.

Figure 5. Lateral view (a) and dorsal view (b) of *Scrapter leovalis* sp. nov female

Male:

Holotype: South Africa, Nieuwoudtville Farm Glen Lyon garden on *Chrysanthemum spec*. S31.23.50 E19.08.28 700m. M. Kuhlmann. 27.viii 2003.

2 Paratypes: South Africa, Nieuwoudtville Farm Glen Lyon, Renosterveld S31.24.03 E19.08.34 700m M. Kuhlmann leg. 10. ix. 2003

Measurements (*n* = 3, 1 holotype and 2 paratypes): head length 1.40-1.60 mm, head width 1.60-1.92 mm, length of clypeus 0.30-0.50 mm, lower interocular distance 1.00-1.30 mm, upper interocular distance 1.00-1.20 mm, length of eye 1.00-1.30 mm, interantennal distance 0.30-0.50 mm, antennocellar distance 0.40-0.60 mm, antennocular distance 0.32-0.40 mm, length or malar area 0.10-0.30 mm, mesoscutum length 1.10-1.30 mm, mesosoma length 2.10-2.43 mm, length of stigma 0.84-0.96 mm, width of stigma 0.20-0.40 mm, length of marginal cell beyond stigma 1.00-1.20 mm, length of marginal cell 1.20-1.30mm, length of free part of marginal cell 0.80-0.95 mm.

Head: White, long, sparse vestiture on face, light brownish or cream white above antenna. Clypeus glabrous, moderate punctation about one puncture diameter apart, no clypeal mediolongitudinal sulcus. Supraclypeus as in female. Scape black, sparse white vestiture. Paraocular area and frons carinate. Facial foveae narrow, sulcus-like. F1-F8 yellow ventrally. F9-F11 brown to black. Eyes converging below. Mandibles as in female. **Mesosoma:** Mesoscutum finely reticulate, shiny, sparse punctation about two to three puncture diameter apart. Scutellum fine reticulate sculptured. Propodeum strongly angulate, basal area roughly sculptured. Pronotal lobe brownish to black. **Metasoma:** T1-T3 brownish, T4-T5 black. T3 and T4 covered in short white vestiture. Foveae on T2. **Legs:** fore tibia and tarsus yellow dorsally. Mesotibia with a yellow spot where joining femur. Metabasitibial plate entire. **Genitalia:** Apex of gonostylus hairy, curved inwards.

Figure 6. Lateral view (a) and dorsal view (b) of *Scrapter leovalis* sp. nov male

Appendix1

Character description

Female and male imago

1. Mandible: (0) all black; (1) basally yellowish to brownish, apically brownish to reddish; (2) basally black, apically brownish to reddish.
2. Frontal line: (0) distinct or easily discernible; (1) indistinct, not easily discernible
3. Compound eyes: (0) converging below; (1) parallel.
4. Vertex seen from front (Michez *et al.*, 2009): (0) flat or weakly convex; (1) convex and elevated well above summits of eyes; (2) concave.
5. Clypeus: (0) greatly convex; (1) weakly convex to flat.
6. Frons: (0) densely to moderately punctate; (1) roughened, almost carinulate.
7. Last flagellar segment (Eardley, 1996): (0) round apically (1) elongate and tapering to acute point.

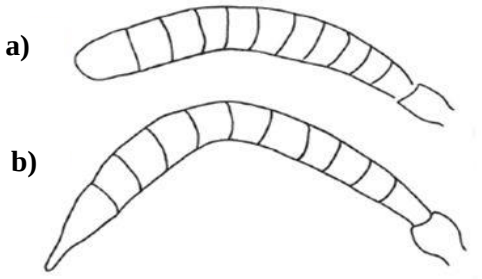

Figure 7. Antenna diagram showing (a) round apical last flagellar segment and (b) elongate and tapering last flagellar segment.

8. Antennal scape (Eardley, 1996): (0) black ; (1) yellow anteriorly and black posteriorly.
9. Paraocular area (Davies and Brothers, 2006): (0) dense to moderate punctation, smooth interspaces, (1) shagreened, near-carinulate sculpture; (2) sparse punctation smooth interspaces.
10. Pronotal lobe (Eardley, 1996): (0) black to red-brown; (1) yellow to cream-white.
11. Propodeal triangle shape (Eardley, 1996): (0) not modified, forming a roughly equilateral triangle, see Fig. 8b (Davies *et al.*, 2005); (1) greatly modified with central area much reduced, triangle forming three acute points, see Fig. 8a (Davies *et al.*, 2005).

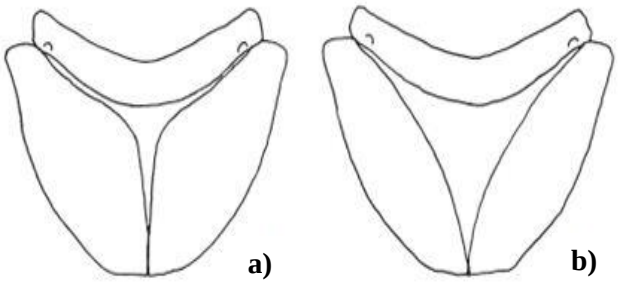

Figure 8. Different propodeal triangle shapes

12. Propodeal triangle sculpture (Eardley, 1996): (0) rough with reticulate sculpture medially; (1) smooth and shiny medially.
13. Propodeum structure: (0) declivitous, see Fig. 9a (Davies *et al.*, 2005); (1) weakly to strongly angulate, see Fig. 9b (Davies *et al.*, 2005).

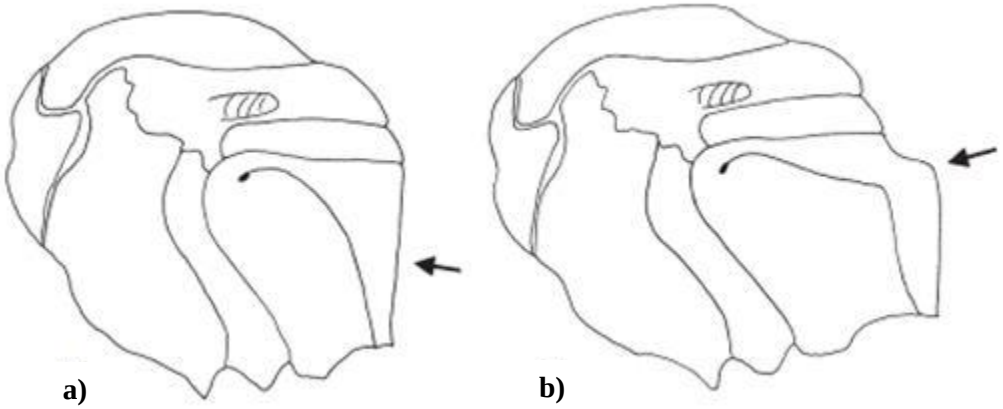

Figure 9. Diagram showing propodeum structures

14. Mesoscutum (Eardley, 1996): (0) densely to moderately hairy; (1) sparsely hairy to naked.
15. Fore wing, stigma (Michez *et al.*, 2009): (0) shorter than first submarginal cell; (1) as long as first submarginal cell.
16. Fore wing, submarginal cells: (0) two, first longer than second; (1) two, first and second of equal length; (2) two, second longer than first.
17. Foveae on the second segment of tergum (T2): (0) present; (1) absent.
18. Claws (Eardley, 1996): (0) basal tooth, see Fig 10a, (1) simple, see Fig10b.

Figure 10. Different claw structures

Male imago

19. Clypeus shape (Eardley, 1996): (0) unmodified, see Fig. 9a (Davies *et al.*, 2005); (1) somewhat reduced, see Fig. 11c (Eardley, 1996); (1) concave ventrally, see Fig. 11b (Davies *et al.*, 2005).

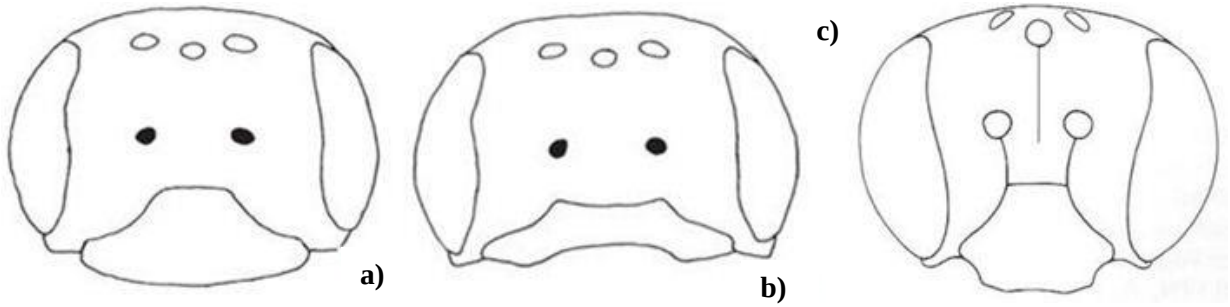

Figure 11. Diagram showing different clypeus structures.

20. Clypeus surface: (0) shiny; (1) dull.
21. Supraclypeal area (Davies and Brothers, 2006): (0) weakly elevated (flattish) with dense punctation, (1) gently to strongly convex with sparse to moderate punctation.
22. Antenna (Eardley, 1996): (0) all or ventrally yellow and black dorsally; (1) all black to dark orange-brown; (2) ventrally yellow except for black apical part of F9 and F10-11.
23. Facial foveae size : (0) narrow; (1) wide.
24. Mesoscutum surface: (0) smooth between punctures; (1) finely reticulate or shagreened between punctures, matt-like in appearance.
25. Mesoscutum structure: (0) generally densely punctate (punctures mostly less than a puncture diameter apart), (1) moderately to sparsely punctate (punctures mostly more than puncture diameter apart).
26. Basitibial plate: (0) entire, see Fig 12a, (1) reduced to a tubercle or absent, see Fig 12b.

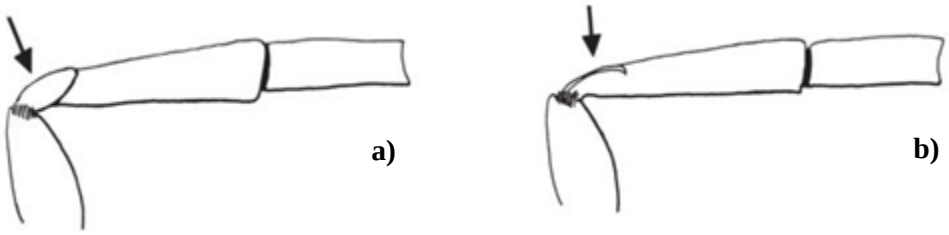

Figure 12. Basitibial plate structures

27. Metatibia (Eardley, 1996): (0) simple, see Fig 13b; (1) tuberculate apically, see Fig 13a.

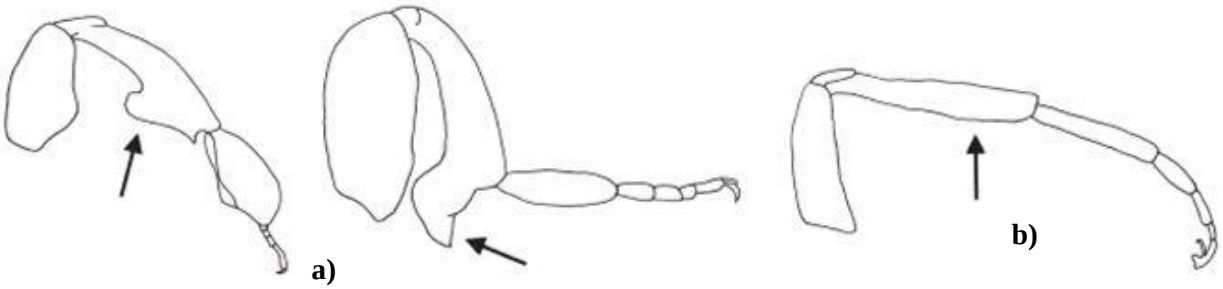

Figure 13. Metatibia structure

28. Fore tibia: (0) all black; (1) yellow anteriorly; (2) all yellow to light orange.
29. Metabasitarsus: (0) flat or slightly swollen, Fig 14b; (1) inflated, Fig 14a.

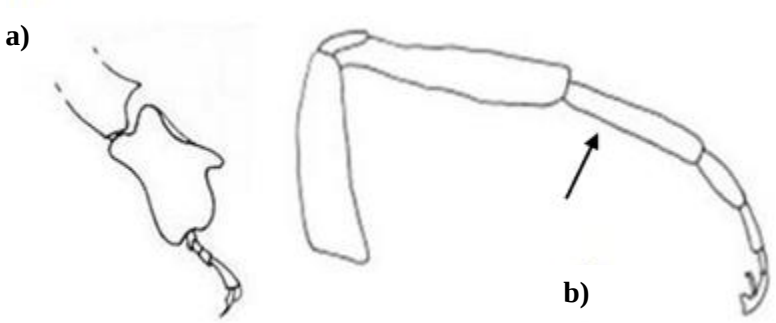

Figure 14. Metabasitarsus structures

30. Mid-tarsus: (0) strongly swollen; (1) not swollen, flat.
31. Gonocoxa: (0) broad and incurved posteriorly; (1) broad posterior end curving outwards; (2) broad narrow posterior end approximately parallel sided.

Figure 15. Different claw structures

Female imago

32. Clypeus mediolongitudinal sulcus (Eardley, 1996): (0) absent, Fig 16b; (1) present, Fig 16a.

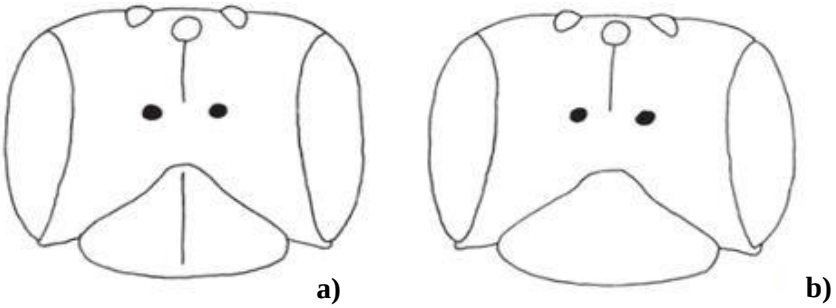

Figure 16. Clypeus with mediolongitudinal sulcus (a) and clypeus lacking mediolongitudinal sulcus

33. Clypeus sculpture: (0) dense punctation; (1) moderately to sparsely punctate; (2) rough almost carinulate sculpture.
34. Facial foveae: (0) shiny; (1) dull.
35. Female mesoscutum surface: (0) smooth between punctures; (1) finely reticulate between punctures, matt-like in appearance.
36. Female mesoscutum structure: (0) generally densely punctate (punctures mostly less than a puncture diameter apart), (1) moderately to sparsely punctate (punctures mostly more than puncture diameter apart).
37. Metasoma colour (Eardley, 1996): (0) largely reddish: (1) black or largely blackish.
38. Metasoma: (0) hairy with clearly visible basal tomentum; (1) largely naked without visible basal tomentum.
39. Wings (Eardley, 1996): (0) infuscated; (1) hyaline.
40. Mid-tibia: (0) black; (1) all yellow or with yellow markings.
41. Mid-tibial spur (Michez *et al.*, 2009): (0) slender; (1) robust, enlarged at base.
42. Basitibial plate: (0) well developed; (1) poorly developed or absent.

43. Scopa (Michez *et al.*, 2009): (0) unicolour; (1) bicolour.

44. Metabasitarsus (Michez *et al.*, 2009): (0) simple; (1) apically divided.

45. Anterior (outer) metatibial spur (Davies *et al.*, 2005): (0) strongly swollen; (1) flat.

46. Prepygidial fimbria (Davies *et al.*, 2005): (0) white or whitish; (1) black.

Appendix 2

Table 3 Data matrix of 46 morphological characters examined on both male and female specimens in 38 taxa including 2 outgroup, 1 putative new species (*S. leovalis* sp. nov) and 35 other Scrapter species.

Taxa	Character codes																																												
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46
L. irroratus	1	1	0	1	1	1	?	?	2	0	0	1	1	1	?	?	?	0	0	0	1	?	0	0	1	0	0	0	0	1	2	0	1	0	0	1	1	0	?	0	0	0	1	?	1	1
L. plumosus	1	1	0	1	1	1	?	?	2	0	0	1	1	1	?	?	?	0	0	0	1	?	0	0	1	0	0	2	0	1	2	0	1	0	0	1	1	0	?	0	0	0	1	?	1	1
S. absonus	1	1	0	2	1	0	0	0	0	0	0	0	1	1	0	1	0	1	0	0	0	0	0	1	1	0	0	1	0	1	0	0	0	1	1	1	1	1	1	0	0	0	0	0	1	0
S. albifumus	2	1	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0	1	1	0	0	2	0	1	1	0	0	2	0	1	1	0	1	0	1	1	1	1	1	0	0	0	0	0	1	0
S. albitarsis	2	0	0	0	1	0	0	0	2	0	0	0	1	1	1	0	0	1	1	0	1	0	0	1	1	0	0	1	0	1	2	0	1	0	0	1	1	1	1	0	0	0	0	0	1	0
S. algoensis	2	0	0	0	1	0	0	0	0	0	0	0	1	0	0	2	0	0	0	0	0	1	0	1	1	0	0	1	0	1	0	0	0	0	1	0	1	1	0	0	0	0	0	0	1	0
S. amplispinatus	2	0	0	0	0	0	0	0	0	0	?	?	1	0	0	1	?	1	0	0	0	1	0	1	0	0	0	0	0	1	0	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
S. amplitarsus	2	1	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	1	0	0	1	1	0	1	0	0	1	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1
S. armatipes	0	1	0	0	1	0	0	0	0	0	0	1	1	1	0	0	0	1	1	0	0	0	0	0	0	0	1	1	1	0	1	0	1	0	0	1	1	1	0	0	0	0	0	0	1	1
S. aureiferus	1	1	0	1	1	0	0	0	0	0	0	0	1	1	0	0	0	1	1	0	?	0	0	0	1	0	1	2	1	1	0	0	1	0	0	1	1	1	1	1	0	0	0	0	1	0
S. avius	1	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	1	0	1	0	0	0	0	1	0	0	2	0	1	1	0	1	0	1	1	1	1	0	1	0	1	0	0	1	0
S. basutorum	1	1	0	0	1	0	0	0	0	1	0	0	1	1	0	1	0	1	0	0	0	0	0	0	1	0	0	1	0	1	1	1	1	1	0	1	1	1	0	1	0	0	0	0	1	1
S. bicolor	2	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	1	1	1	1	1	0	0	1	0	0	0	2	0	1	3	0	1	1	0	1	0	1	0	0	0	1	0	0	1	1
S. caesariatus	0	1	1	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	2	0	1	0	0	1	1	0	0	0	0	1	3	0	1	0	1	1	1	1	1	0	0	0	1	0	0	1
S. calx	2	1	0	0	1	1	0	0	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0	1	0	1	0	1	0	1	0	0	1	1	0	1	1	1	1	1	0	0	0	0	1	0
S. capensis	2	1	0	0	1	1	0	0	0	0	0	0	1	1	0	1	0	1	1	1	1	1	0	1	0	0	0	0	0	1	1	0	1	0	0	1	1	1	0	0	0	1	0	0	1	1
S. catoxys	2	1	0	0	1	0	1	0	0	0	0	1	1	1	0	0	0	1	0	0	0	1	0	0	1	0	0	2	0	1	?	0	1	0	0	1	1	1	1	1	0	0	0	0	1	0
S. chloris	2	0	1	0	1	0	0	0	0	0	0	0	0	1	0	1	1	1	2	1	0	0	0	1	0	0	0	0	0	1	2	0	0	1	1	0	0	1	1	0	0	0	1	0	1	1
S. chrysomastes	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	1	1	1	2	2	2	1	1	1	1	?	?	?	?	?	?	?	?	?	?	?	?	?	?
S. eremanthedon	0	0	?	?	0	1	0	0	1	0	0	0	1	1	?	?	0	0	?	0	1	1	1	0	0	0	0	0	0	1	?	0	0	0	1	1	1	1	?	0	0	1	1	0	1	1
S. erubescens	2	0	0	0	1	0	0	0	0	0	1	0	0	1	0	1	1	0	0	0	1	1	0	1	0	0	0	1	0	1	1	0	1	1	1	0	0	0	0	0	0	0	1	0	1	1
S. flavipes	1	0	0	0	1	0	0	0	0	1	0	0	1	1	0	0	0	1	1	0	0	1	0	1	1	0	0	2	0	1	0	1	0	1	0	0	0	1	1	1	0	0	0	0	1	0
S. flavostictus	1	1	0	0	1	0	0	0	0	1	0	0	1	1	0	0	0	1	1	1	1	0	0	0	1	0	0	2	0	1	0	0	1	0	0	1	1	1	0	0	0	0	0	0	1	0
S. heterodoxus	2	1	0	1	1	1	0	0	1	0	0	0	0	0	0	1	1	0	1	0	0	0	1	1	0	1	1	1	0	1	0	0	0	0	0	0	1	0	1	0	0	0	1	0	1	1
S. leonis	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	1	0	1	0	0	0	1	1	1	1	1	1	1
S. leovalis	2	0	0	0	1	1	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	1	1	0	0	1	0	1	0	0	1	1	1	1	1	1	0	0	0	0	1	0	1	1
S. luridus	1	1	1	0	1	0	0	1	0	0	0	0	1	1	0	0	0	1	0	1	1	0	0	0	0	0	0	2	0	1	1	0	2	1	0	0	0	1	0	1	0	0	0	0	1	0
S. niger	2	0	0	2	1	1	0	0	0	0	0	0	0	1	0	1	0	1	1	1	0	1	1	1	1	1	0	1	0	1	3	0	0	0	0	1	1	1	0	1	0	1	0	0	1	1
S. nitidus	0	1	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	1	0	1	1	0	0	1	1	0	0	0	0	0	1	0	1	1
S. opacus	2	1	0	1	1	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	1	2	1	1	0	0	1	1	1	0	0	0	0	1	0	1	1
S. pallidipennis	2	0	0	0	1	1	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	1	1	1	0	0	0	1	0	1	0	1	0	0	1	1	1	0	1	0	0	0	0	0	1	0
S. pyretus	1	0	0	?	?	0	0	1	0	1	1	0	1	0	?	?	0	0	?	0	0	0	0	0	0	1	0	2	0	1	0	0	1	?	0	0	0	0	?	1	0	1	0	0	?	0
S. pruinosus	1	?	?	?	?	0	0	0	0	0	0	0	1	0	?	?	0	0	?	0	0	0	0	0	0	?	0	2	0	1	0	0	0	0	0	0	0	0	?	1	0	1	0	?	1	0
S. ruficornis	2	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	1	1	0	0	0	0	0	1	0	0	1	0	1	1	1	1	0	0	1	1	1	0	0	0	0	0	0	1	1
S. striatus	2	0	1	0	1	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	1	0	1	0	0	0	1	0	1	2	0	0	1	0	1	1	1	0	0	0	0	0	0	1	0
S. thoracicus	2	?	?	0	0	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	0	1	1	1	1	0	1	2	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
S. tomentum	2	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0
S. whiteheadi	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	2	2	0	0	?	?	?	?	?	?	?	?	?	?	?	?	0	0	1	1	0	0	0	1	1	1	1	1	1	1

Table 4. Data matrix showing outgroup taxa and 4 species with more than 50% missing data which were excluded in the 34 taxa phylogenetic analyses.

Taxa	**Character codes**																																												
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46
L. irroratus	1	1	0	1	1	1	?	?	2	0	0	1	1	1	?	?	?	0	0	0	1	?	0	0	1	0	0	0	0	1	2	0	1	0	0	1	1	0	?	0	0	0	1	?	1	1
L. plumosus	1	1	0	1	1	1	?	?	2	0	0	1	1	1	?	?	?	0	0	0	1	?	0	0	1	0	0	2	0	1	2	0	1	0	0	1	1	0	?	0	0	0	1	?	1	1
S. chrysomastes	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	1	1	1	2	2	2	1	1	1	1	?	?	?	?	?	?	?	?	?	?	?	?	?	?
S. leonis	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	1	0	1	0	0	0	1	1	1	1	1	1	1
S. thoracicus	2	?	?	0	0	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	0	1	1	1	1	0	1	2	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
S. whiteheadi	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	2	2	0	0	?	?	?	?	?	?	?	?	?	?	?	?	0	0	1	1	0	0	0	1	1	1	1	1	1	1

Chapter 3: Molecular phylogeny of *Scrapter*

Introduction

Bees are the largest and most important pollinator group of angiosperms (Danforth *et al.*, 2006a; Danforth *et al.*, 2013). Bees also contribute considerably towards agriculture and the world's economy. One-third of the human diet is derived from fruits, vegetables and nuts which depend on animal-mediated pollination, mostly bee pollination (Klein *et al.*, 2007; Danforth *et al.*, 2013). Having such a crucial contribution ecologically and economically, it is essential we have a clear understanding of bee biodiversity, phylogeny, evolution and diversification (Danforth *et al.*, 2013).

Availability of molecular data and development of new phylogenetic methods have improved understanding of the bee phylogeny in the last fifteen years (Danforth *et al.*, 2013). The use of molecular data is expected to produce highly resolved phylogenies due to the fact that DNA sequences provide far more character-state data than traditional morphological comparisons (Danforth *et al.*, 2013). Therefore, statistically, as more data are examined, a close approximation of patterns of phylogenetic relationships should emerge (Cracraft and Helm-Bychowski, 1991; Miyamoto and Cracraft, 1991; Danforth *et al.*, 2006a; Danforth *et al*., 2013). Also, because DNA stores all inherited attributes of organismal history, direct comparison of DNA data provides the most basic of all data for phylogenetic reconstruction (Cracraft and Helm-Bychowski, 1991). In addition, the congruence among multi-gene data sets is arguably the most reliable indicator of phylogenetic accuracy (Cracraft and Helm-Bychowski, 1991; Miyamoto and Cracraft, 1991; Miyamoto and Fitch, 1995). As part of this study three nuclear protein-coding nDNA loci were examined to reconstruct the phylogenetic history of *Scrapter*.

Nuclear genes have a slower rate of mutation than mitochondrial genes (Springer *et al*., 2011) and are often used in molecular-based phylogenetic analysis of groups of insects because of their large effective sample size (Danforth *et al.*, 1999; 2006a; 2006b; Almeida *et al.*, 2008; Almeida and Danforth, 2009; Springer *et al*., 2011). In many previous studies (e.g. Danforth *et al.*, 1999; 2006a; 2006b; Wiegmann *et al.*, 2000) nuclear genes have been successfully used to recover deep relationships among bee families i.e. to recover Cretaceous-age divergences. For the present study three markers were chosen: Elongation factor F2 copy (EF-1α F2: Hovemann *et*

al., 1998; Almeida *et al.*, 2008); large subunit ribosomal RNA (28SrRNA: Danforth, 2002; Danforth *et al.*, 2004; Michez *et al.*, 2009); and Long wavelength Rhodopsin (Opsin: Sidow, 1992; Mardulyn and Cameron, 1999; Schubert *et al.*, 2000; Michez *et al.*, 2009). The selected molecular markers have been used successfully in the phylogenetic investigations of the family Melittidae (Michez *et al.*, 2009) and Colletidae (Almeida and Danforth, 2009). They have also shown potential in resolving divergences in other insects and arthropods: EF-1α (Danforth and Ji, 1998), 28SrRNA (Cameron and Mardulyn, 2001) and Opsin (Mardulyn and Cameron, 1999).

Previous investigations of the *Scrapter* genus (Eardley, 1996; Davies, 2006; Davies and Brothers, 2006) all used morphological data sets, but morphological data alone could not fully resolve *Scrapter* phylogeny (see Chapter 2). The use of molecular data in this study is expected to shed more light on the classification of *Scrapter* species within the genus and to provide the first phylogeny of this genus based on molecular data. The aim of this chapter is to re-assess the previous morphology-based classification of *Scrapter* using the combined analysis of three molecular data sets and to provide molecular phylogenies of the *Scrapter*.

Materials and methods

Taxon sampling

Molecular data were analyzed from 30 specimens representing 24 species (22 *Scrapter* species and the outgroup taxa *Leioproctus irroratus* and *Leioproctus plumosus* representing Paracolletini a sister group to Scrapterini; Table 5). All 8 morphological species-groups of *Scrapter* (Eardley, 1996) are represented in the molecular data (Table 5). Where possible multiple representative specimens from each species were examined and sequence data from Genbank (Table 5, indicated by *) were combined with the data generated in the present study to verify species identification where morphological characterization was not certain. In two cases (Scrapter_sp1-2013 and Scrapter_sp2-2013; Table 5) species designation was not clear. These specimens resembled none of the known species but possessed key features of *Scrapter* species: the body is elongate, there are two submarginal cells in the fore wing (Davies and Brothers, 2006; Davies *et al.*, 2005; Eardley, 1996; Michener, 2000), the clypeus and paraocular areas lack yellow maculation (Davies and Brothers, 2006). In total, 28 specimens (some species had multiple representatives) were

analyzed for 28SrRNA, 22 taxa for EF-1α and 21 taxa for Opsin. The GenBank accession numbers of the species sourced from GenBank database are listed in Table 5.

Table 5. Taxonomic sampling used in the current study. Species group follows description of Eardley (1996). Species data downloaded from GenBank are indicated by * with accession numbers provided. Sequence data generated in present study are indicated by +, missing data are indicated by -. NCP = Northen Cape Province, WCP = Western Cape Province, N/A = Not applicable.

Species names	Species group	Locality (Provinces)	Coordinates	28SrRNA	EF-1α	Opsin
Outgroup						
*L. irroratus**	Not classified	Australia. NSW, Hilltop. 2.xii.1999	Not available	DQ872765	AY585132	DQ115555
*L. plumosus**	Not classified	Australia. Victoria, Torquay. 19.xi.1999	Not available	DQ872766	AY585133	DQ115556
Ingroup						
*S. algoensis**	Bicolor	NCP. 90 km ENE Springbok. 10.ix.2001	Not available	DQ872771.1	EF032901.1	EF032904.1
S. algoensis	Bicolor	NCP, Springbok.11.ix.2012	29.34 38.2 S, 18.01 10.9 E	+	+	+
S. amplispinatus	Erubescens	NCP, Springbok.12.ix.2012	29.34 38.2 S, 18.01 10.9 E	-	+	+
S. armatipes	Armatipes	NCP, Springbok.11.ix.2012	29.34 38.2 S, 18.01 10.9 E	+	-	+
*S. bicolor**	Bicolor	WCP, Pakhuis pass. 8.ix.2001	Not available	JN566241.1	JN566287.1	-
S. bicolor	Bicolor	NCP, Springbok.11.ix.2012	29.34 38.2 S, 18.01 10.9 E	+	-	+
S. caesariatus	Flavostictus	NCP, Garies.14.ix.2012	30.13 01.4 S, 18.07 04.6 E	+	+	+
S. catoxys	Not classified	NCP, Springbok.12.ix.2012	29.34 38.2 S, 18.01 10.9 E	+	+	+
S. chloris	Bicolor	NCP, Springbok14.ix.2012	29.34 38.2 S, 18.01 10.9 E	+	+	+
*S. chloris**	Bicolor	Not available	Not available	JN603391.1	JN603456.1	-
S. eremanthedon	Not classified	NCP.13.ix.2012	28.49 33.6 S, 17.14 21.6 E	+	+	+
S. erubescens	Erubescens	NCP, Garies.14.ix.2012	30.11 10.9 S, 18.10 04.3 E	+	+	+
S. flavipes	Basutorum	NCP, Springbok.14.ix.2012	29.34 38.2 S, 18.01 10.9 E	-	-	+
*S. heterodoxus**	Heterodoxus	WCP. 31 km S Clanwillian. 7.ix.2001	Not available	DQ872773.1	AY585136.1	DQ115559.1
S. luridus	Bicolor	NCP, Garies.14.ix.2012	30.27 58.9 S, 17.41 45.2 E	+	+	-
S. niger1	Bicolor	NCP, Garies.14.ix.2012	30.11 10.9 S, 18.10 04.3 E	+	+	+
S. niger2	Bicolor	NCP, Springbok.11.ikx.2012	29.34 38.2 S, 18.01 10.9 E	+	-	+
*S. niger**	Bicolor	WCP. 21 km N Hermanus. 28.ix.2001	Not available	DQ872774.1	-	DQ115560.1
*S. nitidus**	Nitidus	Not available	Not available	JN603392.1	JN603457.1	-
S. opacus	Nitidus	NCP, Springbok.14.ix.2012	29.34 38.2 S, 18.01 10.9 E	+	-	+

Table 5. (Continued)

Species names	Species group	Locality (Provinces)	Coordinates	28SrRNA	EF-1α	Opsin
S. opacus	Nitidus	NCP, Springbok12.ix.2012	29.34 38.2 S, 18.01 10.9 E	+	+	+
S. ruficornis	Nitidus	NCP, Springbok.12.ix.2012	29.34 38.2 S, 18.01 10.9 E	+	+	+
Scrapter sp.1-2013	Unknown	NCP, Springbok.12.ix.2012	29.34 38.2 S, 18.01 10.9 E	+	-	-
Scrapter sp. 2-2013	Nitidus	NCP, Garies.14.ix.2012	30.11 10.9 S, 18.10 04.3 E	+	-	+
Scrapter sp. 1 PK 2012	Unknown	Not available	Not available	JN603394.1	JN603458.1	-
*S. striatus**	Striatus	Not available	Not available	JN603393.1	JN603459.1	-
S. tomentum	Bicolor	NCP, Garies.14.ix.2012	30.13 01.4 S, 18.07 04.6 E	+	-	-
S. whiteheadi	Bicolor	NCP, Springbok.12.ix.2012	29.34 38.2 S, 18.01 10.9 E	+	+	-

DNA extraction, PCR and sequencing

DNA extraction was performed using a *Quick-gDNA*™ MiniPrep kit (Zymo Research: The Epigenetics Company). The DNA concentrations were measured using a Thermo scientific Nanodrop 2000 spectrophotometer. The 25µl PCR reactions contained 2µl of 10X DreamTaq™ Buffer, 0.5µl dNTP mix - 2mM each, 0.5µl of each primer (Table 6), 0.1µl DreamTaq™ enzyme, 19.4µl nuclease free water and 2µl DNA (10-35ng/µl). The DreamTaq Green DNA polymerase (Thermo Scientific) comes with a 10X DreamTaq Green Buffer which contains $MgCl_2$ at a concentration of 20 mM. PCR was then performed following the conditions listed in Table 7. To successfully amplify the Opsin and EF-1α gene segments using primers designed in this study an additional 2µl of $MgCl_2$ at concentration of 25mM, was added to PCR reactions. The water volume was adjusted to make up a total volume of 25µl per reaction. PCR products were visualised under UV light in 1.5% agarose gel. The PCR products were run against a 100bp molecular weight marker XIV (Roche). Desired bands were excised from the agarose gel where multiple bands were observed. The bands were chosen based on the expected size of each gene amplified. The excised bands were cleaned using a PCR Clean-up Gel Extraction Kit (NucleoSpin Extract II) and Zymoclean™ Gel DNA Recovery Kit (Zymo Research: The Epigenetics Company). PCR products were sequenced at the Central Analytical Facilities (CAF) at Stellenbosch University using BigDye chemistry (Version 3.1, Applied Biosystems) and analyzed on a 3100 ABI automated sequencer (Applied Biosystems). To check the accuracy of the sequence data and the sequence identity of each amplified fragment, the sequence data were checked by GenBank BlastN searches. All heterozygous sites in the gene sequences were coded using the IUPAC-IUB code.

Table 6. Primer sequences used during the present study.

Primers	Sequences from 5' to 3'	References
28S		
Bel28SFor	5'-AGAGAGAGTTCAAGAGTACGTG-3'	(Belshaw and Quicke, 1997)
A28SFor	5'-CCCCCTGAATTTAAGCATAT-3'	(Ward and Brady, 2003)
Mar28SRev	5'-TAGTTCACCATCTTTCGGGTCCC-3'	(Mardulyn and Whitfield, 1999)
28SD4Rev	5'-GTTACACACTCCTTAGCGGA-3'	(Danforth *et al.*, 2006b)
EF-1α		
HaF2For1	5'-GGGYAAAGGWTCCTTCAARTATGC-3'	(Danforth *et al.*, 1999)
F2rev1	5'-AATCAGCAGCACCTTTAGGTGG-3'	(Danforth *et al.*, 1999)
EF-F	5'-NCANCTGGGCACAGAGAT-3'	Designed in this study
EF-R	5'-ATGTGACGATCATACCCGG-3'	Designed in this study

Table 6. (Continued)

Primers	Sequences from 5' to 3'	References
Opsin		
OpsinF3mod	5'-TTCGAYAGATACAACGTRATCGTNAARGG-3'	(Danforth *et al.*, 2004)
OpsinRevmod	5'-ATANGGNGTCCANGCCATGAACCA-3'	(Danforth *et al.*, 2004)
Ops-F	5'-ATCTGCTAAGCCACTG-3'	Designed in this study
Ops-R	5'-CGATGGGAATWTCTGG-3	Designed in this study

Table 7. Polymerase chain reaction conditions for each primer pairs used during the present study.

Primer pair		PCR Conditions (35 cycles)	
28S			
A28SFor/Mar28SRev	94ºC/ 1:30	94ºC/ 1:00 ; 58ºC/1:00; 72ºC/1:00	72ºC/10:00
D2BF/D3AR	95ºC/ 3:00	95ºC/ 0:30 ; 58ºC/ 0:30; 72ºC/1:30	72ºC/10:00
EF-1α			
HaF2For1/F2rev1	94ºC/ 1:30	94ºC/ 1:00 ; 54-6ºC/1:00; 72ºC/1:00	72ºC/7:00
Elo F-F/Elo F-R	94ºC/ 1:30	94ºC/ 1:00 ; 55ºC/1:00; 72ºC/1:00	72ºC/7:00
Opsin			
OpsinF3mod/OpsinRevmod	94ºC/ 2:00	94ºC/ 1:00 52-4ºC/1:00; 72ºC/1:00	72ºC/7:00
Ops-F/Ops-R	94ºC/ 1:30	94ºC/ 1:00; 52-6ºC/1:00; 72ºC/1:00	72ºC/7:00

Phylogenetic methods

Gene sequences were analyzed separately and combined for the supermatrix analysis using parsimony, maximum likelihood and Bayesian analyses. Sequence alignments were generated using ClustalX 2.1 (Larkin *et al.*, 2007) and edited in BioEdit Sequence Alignment Editor package (Hall, 1999). Unweighted parsimony analyses were conducted using Tree Bisection Reconnection (TBR) in MEGA v5 (Tamura *et al.*, 2011). I performed 1000 bootstrap replicates with 10 random additions per replicate.

Using the Akaike information criteria (AIC: Akaike, 1974) in jModelTest v2.1.4 (Darriba *et al.*, 2012) the best-fit model for each gene was estimated. The Hasegawa substitution model (Hasegawa *et al.*, 1985)with gamma distributed rates (HKY+G) was indicated to best fit the opsin gene data set, the Tamura-Nei (TrN+G) model (Tamura and Nei, 1993) was indicated to be suitable for the EF-1α data set and the Transition (TIM3+G) model (Posada, 2003) was indicated to best fit the 28SrRNA data set (Table 8). For each gene data set two maximum likelihood analyses were performed in GARLI v0.951 (Zwickl, 2006) using the best-fit models estimated (Table 8). The first run consisted of five replicate searches without bootstrap replicates to obtain

the most likely tree and the second run was performed with 100 bootstrap replicates to examine branch support. A 50% majority rule consensus tree was constructed using the Consense program in the PHYLIP package (Felsenstein, 2004).

Bayesian inference (BI) was carried out using MrBayes v3.2 (Huelsenbeck and Ronquist, 2003). Substitution models for different genes were applied (Table 8). Analyses consisted of running four simultaneous chains for 10 million generations with trees sampled every 1000th generation. Convergence was assessed using the average standard deviation of split frequencies and inspection of trace plots, density curves and effective sample sizes using Tracer v1.5 (Rambaut and Drummond, 2009). Effective sample size values (ESS: Kass *et al.*, 1998) give a general idea of the balance between the length of the chain and the posterior sampling frequency during the run. An ESS of 200 and above means that sufficient sampling has occurred (Rambaut and Drummond, 2009). The first 2500 trees were removed as burnin from tree files. Four separate runs were performed to verify that the resulting tree topologies were not biased by different starting points. A 50% majority rule consensus tree was constructed using the Consense program in the PHYLIP package. For both Bayesian and maximum likelihood analyses data were partitioned by gene and for each gene partition the appropriate substitution model was applied (Table 8) and analyzed simultaneously.

Table 8. Substitution models estimated in jModelTest v2.1.4 using Akaike information criteria (AIC) which were used for the three nuclear markers.

Gene	Model name	Lset	Rates	Rate matrix	State frequencies
28SrRNA	TIM3+I+G, (Posada, 2003)	nst=6	Gamma	(0 1 2 2 3 0)	Estimate
EF-1α	TrN+G, (Tamura and Nei, 1993)	nst=6	Gamma	(0 1 0 0 2 0)	Estimate
Opsin	HKY+G (Hasegawa et al., 1985)	nst=2	Invgamma	2rate	Equal

Results

The final combined data set consisted of 2269 aligned nucleotide sites. The 28S data included 29 taxa which amounted to 970 aligned nucleotide sites with 103 parsimony informative sites. The EF-1α data had 21 taxa and 593 aligned nucleotide sites which produced 74 parsimony informative sites. And the Opsin data included 22 taxa and 706 aligned nucleotide sites which showed 223 parsimony informative sites. The summarized information about the number of

aligned nucleotides, number of parsimony informative characters and measure of data fit (consistency and retention index) is listed in Table 9. Of the three nuclear genes used the Opsin gene appeared to be most rapidly evolving with 46% variable sites compared to EF-1α, 20% and 28SrRNA, 16% (Table 9). The 28SrRNA gene was the most conserved marker out of three nuclear markers with 82.5% conserved sites in the sequence. And EF-1α gene was also conservative with 79% constant characters in the alignment; Table 9.

Phylogenetic analyses

Bootstrap consensus trees were constructed for the parsimony analysis of each gene and are shown below (Figure 17). Measure of fit was estimated in MEGA v5 (Tamura *et al.*, 2011) and all three nuclear markers showed high consistency (CI) and retention indexes (RI) values; (28SrRNA showed CI=0.796 and RI=0.859), EF-1α showed CI=0.795 and RI=0.856 and lastly Opsin had CI= 0.872 and RI= 0.879). These values indicate that the alignment analyzed contain few homoplasious characters (Farris 1969).

Table 9. Details of the three molecular markers included in the present study. Measure of data fit (consistency index and retention index) were estimated in MEGA v5.

	28SrRNA	**EF-1α**	**Opsin**	**Supermatrix**
No. of aligned sites	970	593	706	2269
No. of conserved sites	800	468	379	1647
No. of variable sites	156	119	322	597
No. of parsimony informative sites	103	74	223	407
No. of most parsimonious trees	5	5	4	2
Tree length	147	88	218	902
Consistency index	0.796	0.795	0.872	0.829
Retention index	0.859	0.856	0.879	0.819
Parsimony informative consistency index	0.737	0.739	0.827	0.750
Parsimony informative retention index	0.859	0.856	0.879	0.819
Parsimony informative composite index	0.663	0.633	0.728	0.610

Maximum likelihood trees (not shown) were identical to trees obtained in Bayesian analyses, therefore maximum likelihood bootstrap values were then incorporated on the Bayesian trees.

Posterior probability values less than 0.80 and bootstrap values less than 50% are not shown on branches in Figures 17, and 18. The number of nodes resolved by each gene independently differed with the Opsin gene data recovering the most resolved and well-supported phylogeny (Figures 17c and 18c) presumably because of the high number of variable and parsimony-informative sites (Table 9) compared to other genes analyzed.

The high bootstrap support observed in the parsimony (Figure 17a) and maximum likelihood (Figure 18a) results and the high posterior probabilities (Figure 18a) of the 28SrRNA are attributable to most species having multiple representatives which make relationship between the two branches of the same species moderately well supported e.g. *S. niger* cluster (three *S. niger* specimens received 72% parsimony bootstrap support; 73% maximum likelihood bootstrap support and 1.00 Bayesian posterior probability support). There was support for a consistent close relationship between *S. caesariatus* and *S. striatus* in Figure 17a (99%) and Figure 18a (97% maximum likelihood bootstrap and 1.00 Bayesian posterior probability).

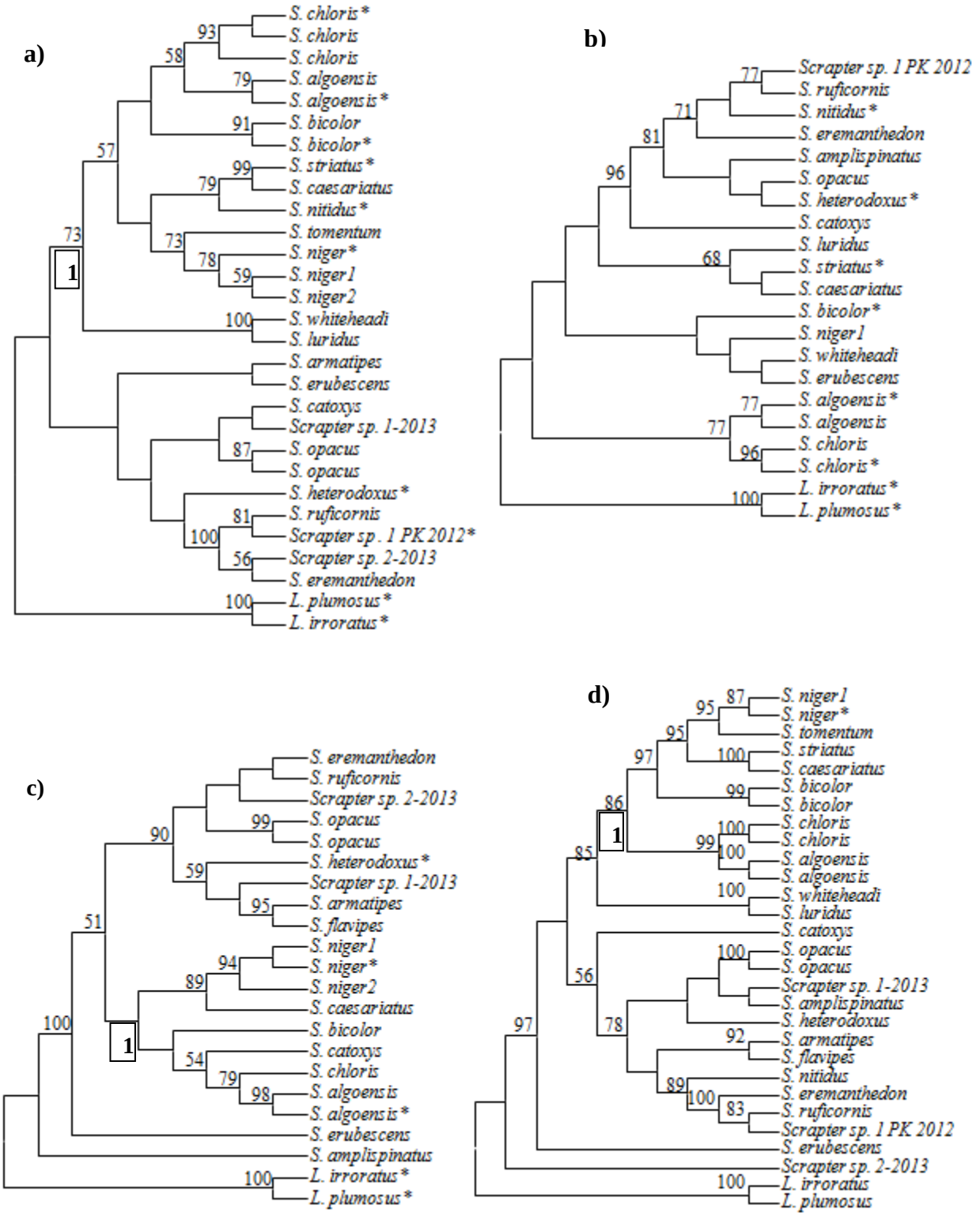

Figure 17 The unweighted parsimony bootstrap consensus trees of (a) 28SrRNA, CI=0.796, RI=0.859; (b) EF-1α, CI=0.795, RI=0.856; (c) Opsin gene, CI=0.872; RI=0.879 and (d) supermatrix analysis, CI=0.829, RI=0.819. The values above the branches are bootstrap support values.

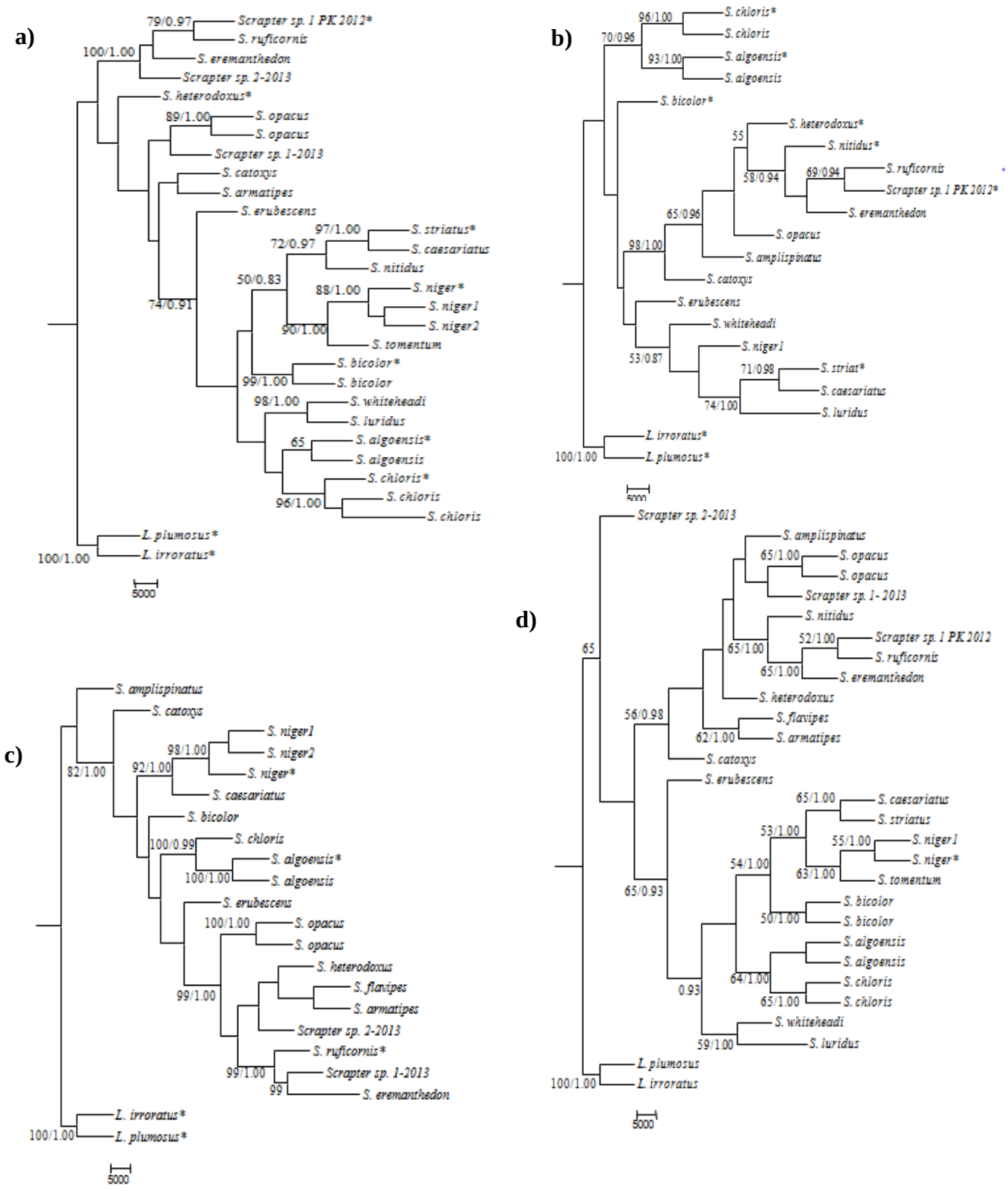

Figure 18. Bayesian trees with Bayesian posterior probability values and maximum likelihood bootstrap values annotated onto the trees; (a) 28SrRNA; (b) EF-1α; (c) Opsin and (d) supermatrix analysis. Values on the right side of the solidus (/) are posterior probabilities from the Bayesian analysis and on the left side are the bootstrap values from the maximum likelihood analysis.

The combined analyses of all available molecular data produced improved nodal posterior probability support of the phylogeny (Figure18d). The total of 28 sequences representing 24 species were included in the supermatrix analysis and the tree was well resolved with accepted bootstrap and posterior probability support values at many nodes (Figure 18d). In the maximum likelihood and Bayesian supermatrix topology, *S. erubescens* (1.00 posterior probability and 50% bootstrap support) is placed at the base of *Scrapter*. The *S. caesariatus* + *S. striatus* clade received 99% bootstrap support (Figure 17d), 97% bootstrap and 1.00 posterior probability (Figure 18d). The *S. chloris* + *S. algoensis* clade received 99% bootstrap support (Figure 17d), 64% bootstrap and 1.00 posterior probability (Figure 18d). The *S. striatus* + *S. caesariatus* clade received 100% bootstrap support (Figure 17d), 65% bootstrap support and 1.00 posterior probability (Figure 18d). The *S. niger* + *S. tomentum* clade received 95% bootstrap support (Figure 17d), 63% bootstrap support and 1.00 posterior probability (Figure 18d).

Discussion

Previous classifications of *Scrapter* relied heavily on morphological characters (Davies, 2006; Eardley, 1996; Michener, 2007); this study presents the results of the analysis of sequence data from three nuclear markers. The molecular data does provide some support for the recognition of some of the associations suggested by Eardley (1996) and Michener (2007). Based on morphological data the *bicolor* group consists of ten species (*S. bicolor, S. whiteheadi, S. niger, S. leonis, S. thoracicus, S. tomentum, S. algoensis, S. albifumus, S. chloris* and *S. luridus*). In this study, seven of these species were included in the molecular analyses and formed a group in Figures 17a, c, d and Figures 18a and d. There was a close relationship between this group ("*bicolor* species group") and the *S. caesariatus* and *S. striatus* clades. The *S. caesariatus* falls under the *S. flavostictus*-group and *S. striatus* under the *S. striatus*-group (Eardley, 1996). Morphologically, these three species-groups closely resemble each other and are characterized by several morphological characters including the posteromedially pointed male S7 and the lack of a distinct mediolongitudinal clypeal sulcus (Eardley, 1996). Two species (*S. nitidus* and *S. ruficornis*) belonging to the *nitidus*-group (Eardley, 1996) were also recovered in one clade Figure 17b and 18b. Michener's (2007) classification is mostly based on female specimens which may not be accurate, for example the *S. flavipes* female has a wide facial foveae (Michener 2007) which would cluster this species in Group 1 but the facial foveae of the male is narrow which is a trait diagnostic for Group 2 species. This creates confusion among species

groupings based on morphology. Another example is *S. albifumus* male which has narrow facial foveae (thus assigned to Group 2) but the thoracic sculpturing is finely reticulate and the basitibial plate forms two carinae. The latter two traits are observed in the Group 1 (Michener, 2007).

As this study was the first to examine the evolution of the *Scrapter* genus using molecular data, it can be used as a reference study in future studies investigating *Scrapter*. Further molecular work on the phylogeny and species identification, using molecular barcodes, is still needed to contribute additionally to our understanding of *Scrapter* taxonomy and phylogeny, as there were limited number of species in the current study.. Additional species sampling and molecular character sampling (mitochondrial DNA) are needed to revise the taxonomy of the group. Nonetheless, the phylogenetic trees based on the molecular data (Figures 17 and 18) presented in the present study suggest that the chosen morphological characters alone are not adequate for resolving classification of *Scrapter*.

Chapter 4: Supermatrix Analysis: combining morphological and molecular data

Introduction

Depending on the available data, large phylogenetic trees could be constructed using many independent data sets such as molecular sequences, morphology, behaviour, genomic arrangements and other traits (De Queiroz and Gatesy, 2006). These separate data sets can be combined and analyzed to produce a single phylogenetic hypothesis. Two methods have been proposed to achieve this: the supertree and the supermatrix (De Queiroz and Gatesy, 2006). The supertree method involves analyzing different data sets independently. The trees inferred from each data set are then used as input in a second analysis which produces the final supertree phylogeny (De Queiroz and Gatesy, 2006). However, some character information is lost when analyzing trees instead of analyzing the data directly (De Queiroz and Gatesy, 2006).

In the supermatrix method all data sets are combined into a single data matrix and analyzed simultaneously (De Queiroz and Gatesy, 2006). Using a supermatrix approach is advantageous, over the supertree method, for several reasons. First, in a supermatrix analysis character evidence is used more fully in estimating the tree (De Queiroz and Gatesy, 2006). An important aspect of this full, direct use of data is that the phylogenetic signal in a supermatrix can be greatly improved from that which is apparent in the trees from separate analyses, presumably because the combined analysis in the supermatrix enables the signal to assert itself more strongly over noise (Barrett *et al.*, 1991). In contrast, including data which is not phylogenetically informative (i.e. which is convergent, homoplasy), may increase noise and actually reduce resolution in the supermatrix (Barrett *et al.*, 1991). Combining different datasets (such as morphological and molecular) will most likely produce more reliable end results and allow for close scrutiny of conflicts between different datasets (e.g. Simonsen *et al.*, 2006; Warren *et al.*, 2009). In the present study we evaluated the utility of the supermatrix analysis by simultaneously analyzing the molecular data from three nuclear markers (EF-1α, Opsin and 28SrRNA; Chapter 3) and morphological data set including 46 characters (Chapter 2). Because phylogenetic resolution can be reduced by including characters that are not tracing the phylogenetic history of the group, I also examined where morphological and molecular evidence agree and where they seem to be in conflict.

Materials and methods

Supermatrix analyses can recover hidden support, that is high support for a clade in a combined analysis relative to that recovered by separate analyses of the individual data sets (Barrett *et al.*, 1991; De Queiroz and Gatesy, 2006). But extensive missing data can bias parameters that are based on summations of all characters, for example branch lengths, nucleotide composition biases and corrected distances between taxa (Gatesy *et al.*, 2002). This is relevant because several species analyzed in Chapter 2 were not available for the molecular study. To limit the levels of bias through missing data, species had to have at least one gene data set and morphological data to be included in this analysis as shown in Table 10. Therefore only 21 species were included in this study.

Table 10. The list of taxa and different data sets incorporated in the supermatrix analysis**. (**+) = data present for that particular data set; (-) = data absent for that particular data set. Taxa shaded in blue only have morphological data.

Taxa	28SrRNA	Ef-1α	Opsin	Morphology
L. irroratus	+	+	+	+
L. plumosus	+	+	+	+
S. algoensis	+	+	+	+
S. amplispinatus	-	+	+	+
S. armatipes	+	-	+	+
S. bicolor	+	+	+	+
S. caesariatus	+	+	+	+
S. catoxys	+	+	+	+
S. chloris	+	+	+	+
S. eremanthedon	+	+	+	+
S. erubescens	+	+	+	+
S. flavipes	-	-	+	+
S. heterodoxus	+	+	+	+
S. luridus	+	+	-	+
S. niger	+	+	+	+
S. nitidus	+	+	+	+

Table 10. (Continued)

Taxa	28SrRNA	Ef-1α	Opsin	Morphology
S. opacus	+	+	+	+
S. ruficornis	+	+	+	+
S. striatus	+	+	-	+
S. tomentum	+	-	-	+
S. whiteheadi	+	+	-	+

The supermatrix analysis was conducted using Bayesian and maximum likelihood methods. Parsimony analysis was not conducted for this chapter for the following reason. Parsimony is not a model based analyses unlike Bayesian and maximum likelihood analyses. With model based analyses one is able to partition data sets accordingly, therefore enabling in-depth analysis of the different data sets by applying different evolutionary models on each data set. The Bayesian analyses were conducted in MrBayes 3.1.2 (Huelsenbeck and Ronquist, 2003). A partitioned approach was taken with each nuclear gene and the molecular data assigned to a different partition. Models applied for each gene partition are the same as those in Table 8 (Chapter 3) and the Mkv model (Chapter 2) was applied to the morphological data partition. Analyses consisted of running four simultaneous chains for 10 million generations. Four separate runs were performed to verify that the resulting tree topologies were not biased due to different starting points of the MCMC (Markov chain Monte Carlo) chain (Danforth *et al.*, 2006a). Trees were sampled every 1000th generation. After each Bayesian run, convergence was assessed in Tracer v1.5 (Rambaut and Drummond, 2009). The stationary state of the MrBayes run was evaluated by inspecting the trace plots, density curves and the effective sample size values (ESS: Kass *et al.*, 1998).

For maximum likelihood analysis I used Garli v0.951 (Zwickl, 2006) with two runs. The first run consisted of five searches without bootstrap replicates. In the second run 100 bootstrap replicates were constructed to assess branch support. Bayesian and maximum likelihood consensus trees were constructed using the Consense programme in the PHYLIP package (Felsenstein, 2004). Phylogenies were visualized and edited in FigTree v1.3.1 (Rambaut, 2009).

To examine if the morphological characters were tracing the molecular phylogeny Consistency Index (CI) values were calculated in Mesquite (Maddison and Maddison, 2001) for each character using the molecular phylogeny. A graph was plotted showing CI for each morphological character in Figure 12. Each morphological character was traced on to the molecular phylogeny (combined analysis of all three nuclear genes Chapter 3, Figure 18d). The CI values (Farris 1969) range from 0 to 1 and provide a measure of how data fits a phylogeny. Values close to zero indicate high levels of homoplasy and in this study would suggest that a morphological character was in conflict with the molecular phylogeny. In contrast CI values close to one suggest that a morphological character is in agreement with the molecular phylogeny.

Results and discussion

Bayesian analysis recovered a well supported phylogenetic tree showing most branches with posterior probability support above 0.90 (Figure 19). A combined analysis of three genes and morphological characters recovered the same species classification within *Scrapter* as have been shown in Figure 18d, only with higher posterior probability support.

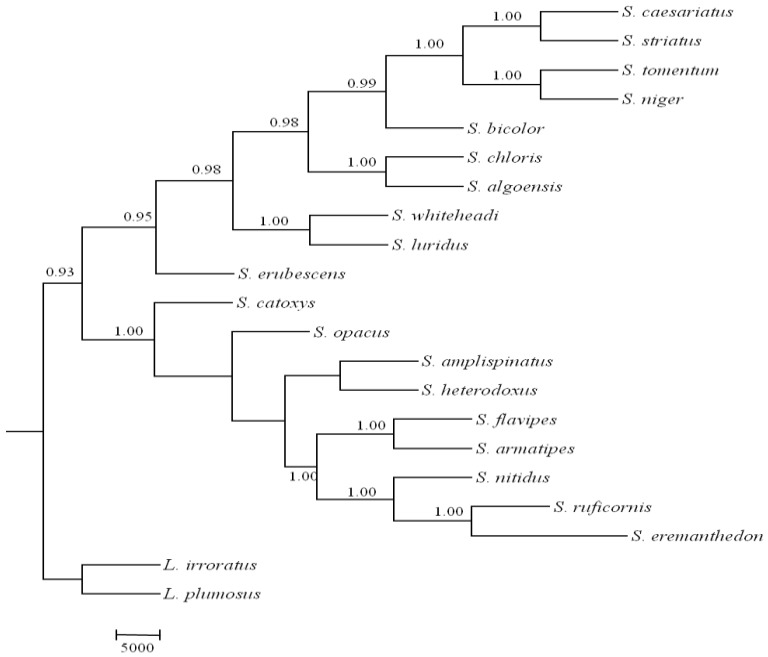

Figure 19. Bayesian combined analysis of the 28SrRNA gene, EF-1α gene, Opsin gene and morphological characters of 21 *Scrapter* species. Branch support is measured by posterior probabilities and posterior probability < 0.80 is not shown.

The (*S. caesariatus* + *S. striatus*) clade received high (1.00) posterior probability support (Figure 19) and bootstrap support (99%) bootstrap in Figure 20. This relationship was also supported by separate datasets in molecular analysis (Chapter 3, Figures 17 and 18) but when only the morphological data set was analyzed (Chapter 2; Figures 3 and 4) these species did not form a monophyletic lineage; in Figure 2 the phylogenies were not resolved.

When only morphological data set (Figures 3 and 4 in Chapter 2) was analyzed both *S. eremanthedon* and *S. ruficornis* were recovered in different positions in the phylogenies. However, in the bayesian and maximum likelihood combined analyses *S. eremanthedon* and *S. ruficornis* association was highly supported in Figures 19 and 20 with 1.00 posterior probability and 100% bootstrap support, respectively.

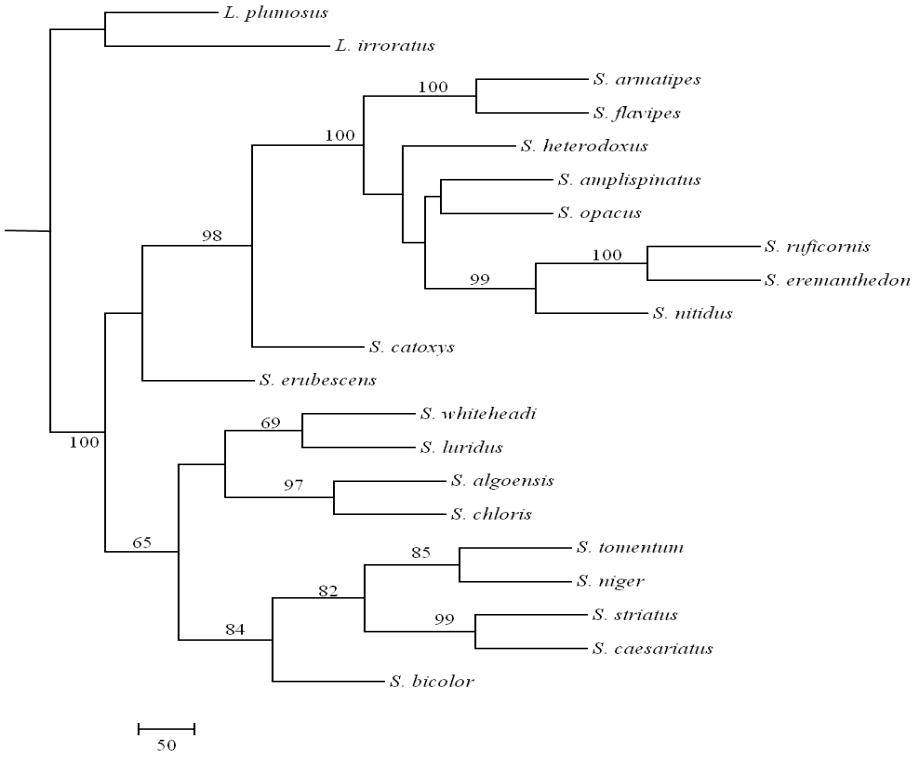

Figure 20. Maximum likelihood analyses of the combined data viz. 28SrRNA gene, EF-1α gene, opsin gene and morphological charaters of 21 *Scrapter* species. Bootstrap support values less than 50% are not shown.

The combined data analysis generally supported relationships that were recovered by the analysis of molecular data in Chapter 3. This was expected given that the majority of characters in the supermatrix are molecular and the phylogenetic signal of larger data can overwhelm that of

smaller data sets (46 morphological characters) in a combined analysis (Miyamoto, 1985). Several authors (Lamboy, 1994; Hedges and Maxson, 1996; Givinish and Systma, 1997) have disapproved the utility of morphological data on the basis that they are more prone to convergence than molecular data and therefore likely to provide misleading results. Interpretation of the supermatrix results requires ultimate caution due to uncertainty of the balance between the separate data sets used. The observed results could be because of the hidden support factor explained by De Queiroz and Gatesy (2006) i.e. the high branch support observed in the supermatrix analysis relative to that observed in individual data set separate analyses; the alternative would be the molecular data biasing the analysis because of the larger number of characters contained in the molecular data compared to the morphological data.

Out of 46 morphological characters examined only 7 morphological characters had a consistency index of one (CI = 1) as shown in Figure 21. These characters are, (i) character 7 = shape of the last flagellomere in both sexes, (ii) character 8 = color of the antennal scape in both sexes, (iii) character 10 = colour of the pronotal lobe in both sexes, (iv) character 11 = shape of the propodeal triangle in both sexes, (v) character 29 = structure of the male metabasitarsus, (vi) character 30 = structure of the male mid-tarsus and (vii) character 45 = shape of the female metatibial spur.

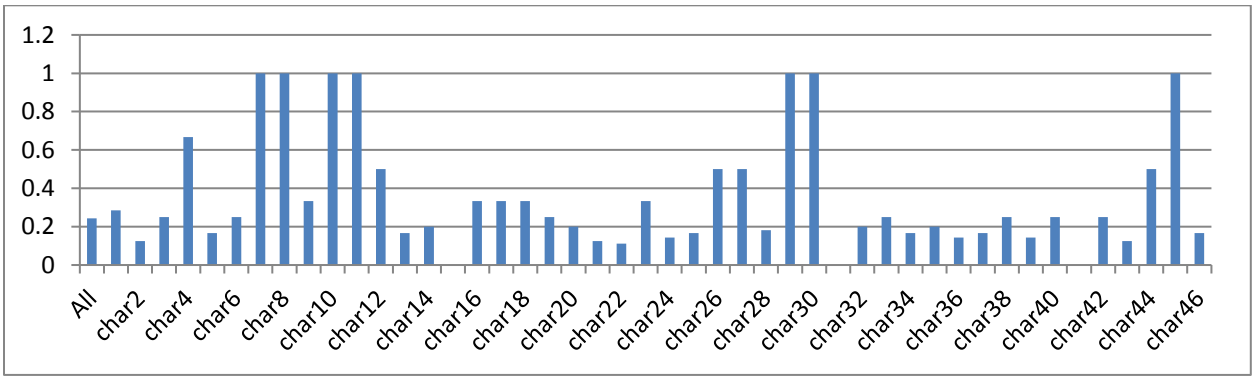

Figure 21. Phylogenetic signal present in the morphological characters estimated by CI values.

Characters 7, 8, 10, 11, 29 and 30 are autapomorphic characters and although diagnostic for specific species they provide no information on the evolutionary relationships among species. Results based on molecular data in this study best define species relationship of *Scrapter*. Molecular analysis of individual data sets and combined data sets produced highly supported

phylogenies with moderately high consistency index 0.83 and retention index 0.87 compared to phylogenies produced from the analysis of morphological data set. From the Figure 21 we can conclude that molecular data examined better estimates the *Scrapter* classification compared to the morphological characters that were included in this investigation.

Chapter 5: General discussion and conclusion

The genus *Scrapter* is a group of bees endemic to the southern African region. Previous cladistic investigations of *Scrapter* have relied only on the analysis of morphological characters (Davies, 2006; Michener, 2007). These studies selected different suites of characters. This study aimed to test previous morphological classifications of *Scrapter* through combined analysis of molecular sequences and morphological data within a phylogenetic framework. In Chapter 2 I reexamined the morphological characters used in previous studies and analyzed a matrix including 46 morphological characters from 38 and 34 ingroup taxa. This was done to observe if the missing data could negatively affect the phylogenetic assumption of *Scrapter* classification. The 38 taxa data set consisted of four species (*S. whiteheadi, S. thoracicus*, *S. leonis* and *S. chrysomastes*) with more than 50% missing data. These four species were excluded in the 34 taxa data set. The effect of missing data was not witnessed as the phylogenies in Chapter 2 did not vary in branch support or clade resolution, rather the limited number and poor fitness of morphological characters coded contributed to poor resolution of phylogenies in Figures 2 and poor branch support in Figures 3 and 4.

The phylogenetic trees constructed using parsimony had low consistency and retention indexes, which suggest that the morphological characters have a high level of homoplasy and that the phylogenies based on these morphological characters alone may not be reliable. The analysis of morphological characters by model-based analysis is controversial because it is not clear how model fit will affect the final topology (Givnish and Sytsma, 1997), and in some cases the positions of the clades in the maximum likelihood and the Bayesian inference trees recovered in my study varied. Despite these limitations, the model-based analyses did recover similar clades. The *S. nitidus* clade comprising of *S. nitidus, S. opacus, S. eremanthedon* and *S. caesariatus* was observed in all three analyses. The most significantly supported nodes, the *S. pruinosus* and *S. pyretus* clade and *S. eremanthedon* and *S. caesariatus* clade were also well supported in all three analyses. My analysis was also able to confirm the specific status of *S. leovalis* sp. nov although the position of this taxon within *Scrapter* was unresolved. The poor bootstrap and posterior probability support observed in the morphological topology is partially because of the limited number of morphological characters chosen and the susceptibility to homoplasy of morphological characters (Lamboy, 1994; Hedges and Maxson, 1996; Givnish and Sytsma,

1997). Given the high levels of homoplasy and the low support for the branches observed in the morphology-based phylogeny, in Chapter 3 I used sequence data sets from three nuclear genes which were the Elongation factor (EF-1α), large subunit ribosomal RNA (28SrRNA) and Long wavelength Rhodopsin (Opsin) to produce for the first time molecular phylogenies looking at species classification among *Scrapter* genus. Molecular data recovered clades with high bootstrap and posterior probability support compared to the morphological data analyses. The recovery of *S. bicolor* group was consistent among the three genes examined and among three phylogenetic analysis carried out which were parsimony, maximum likelihood and bayesian. With an acceptable success of molecular data analyses over the morphological data analyses it was concluded that molecular data best classified *Scrapter* species.

Simultaneous analyses of molecular and morphological data

Supermatrix approaches are not only popular for producing robust phylogenies but also for easily accommodating different classes of data sensibly (De Queiroz and Gatesy, 2006). In Chapter 4 I evaluated an analysis of combined morphological and molecular data sets. This simultaneous analysis (Chapter 4 Figures 19 and 20) recovered well supported trees. It also yielded highly incongruent results compared to the morphological results, but showed similarities in comparison to the molecular data results in Chapter 3 as most identical clades were recovered and highly supported. Molecular data sets are known to have greater potential for conveying phylogenetic information than morphological data sets because of the number of characters taken into consideration when conducting a phylogenetic analysis (De Queiroz and Gatesy, 2006).

The conflict between the morphology and molecular data observed in this study is not new, it has also been observed in other studies (e.g. Danforth *et al,*. 2006b combining five genes with 109 morphological characters; Michez *et al.,* 2009 using five genes and 68 morphological characters; and Koch, 2010 combining morphological measurements with the COI gene). In conclusion the study successfully classified *Scrapter* species with the aid of simultaneous analysis of morphological and molecular data. The coded morphological data alone did not produce reliable phylogenies as they suffered from low consistency and retention indices, poor bootstrap and posterior probability supports and also poor resolution in the parsimony analysis. To overcome these factors in future studies, it is recommended to increase the amount of

morphological characters and include as many *Scrapter* species as possible in the investigation. This study provided a platform for future coming studies investigating classification of *Scrapter* species, as such, the outcome of this study is essential to understanding evolutionary history of the *Scrapter* genus.

References

AKAIKE, H., 1974. A new look at the statistical model identification. *Automatic Control, IEEE Transactions* **19:** 716-723.

ALEXANDER, B. A, and C. D. MICHENR, 1995. Phylogenetic studies of the families of short-tongued bees (Hymenoptera: Apoidea). *The University of Kansas Science Bulletin* **50** (11)**:** 377-424.

ALMEIDA, E. A. B., and B. N. DANFORTH, 2009. Phylogeny of colletid bees (Hymenoptera: Colletidae) inferred from four nuclear genes. *Molecular Phylogenetics and Evolution* **50:** 290-309.

ALMEIDA, E. A. B., L. PACKER and B. N. DANFORTH, 2008. Phylogeny of the Xeromelissinae (Hymenoptera: Colletidae) based upon morphology and molecules. *Apidologie* **39:** 75-85.

ANDERSON, J. S., 2001. The phylogenetic trunk: Maximal inclusion of taxa with missing data in an analysis of Lepospondyli (Vertebrata, Tetrapoda). *Systematic Biology* **50:** 170-193.

ARMBRUSTER, W. S., 1984. The role of resin in angiosperm pollination: ecological and chemical considerations. *American Journal of Botany***:** 1149-1160.

ASCHER, J., and J. PICKERING, 2012. Bee species guide (Hymenoptera: Apoidea: Anthophila). Discover Life, http://www.discoverlife.org/mp/20q?guide=Apoidea_species.

ASHMAN, T. L., T. M. KNIGHT, J. A. STEETS, P. AMARASEKARE, M. BURD, D. R. CAMPBELL, M. R. DUDASH, M. O. JOHNSTON, S. J. MAZER, R. J. MITCHELL, M. T. MORGAN and W. G. WILSON. 2004. Pollen limitation of plant reproduction: ecological and evolutionary causes and consequences. *Ecology* **85:** 2408-2421.

BARRETT, M., M. J. DONOGHUE and E. SOBER, 1991. Against consensus. *Systematic Zoology* **40:** 486-493.

BARTOMEUS, I., J. S. ASCHER, D. WAGNER, B. N. DANFORTH, S. COLLA, S. KORNBLUTH and R. WINFREE. 2011. Climate-associated phenological advances in bee pollinators and bee-pollinated plants. *Proceedings of the National Academy of Sciences of the USA* **108:** 20645-20649.

BELSHAW, R., and D. L. QUICKE, 1997. A molecular phylogeny of the Aphidiinae (Hymenoptera: Braconidae). *Molecular Phylogenetics and Evolution* **7:** 281-293.

BUCHMANN, S. L., 1987. The ecology of oil flowers and their bees. *Annual Review of Ecology and Systematics* **18:** 343-369.

BUCHMANN, S. L., and G. P. NABHAN, 1996. The Forgotten Pollinators. Island Press/Shearwater Books, Washington, D.C/Covelo, Carlifornia.

CAMERON, S. A., and P. MARDULYN, 2001. Multiple molecular data sets suggest independent origins of highly eusocial behavior in bees (Hymenoptera: Apinae). *Systematic Biology* **50:** 194-214.

COWLING, R. M., and C. HILTON-TAYLOR, 1994. Patterns of plant diversity and endemism in southern Africa: an overview: Huntley, B. J., ed., *Botanical diversity in southern Africa.*, pp. 31-52, Pretoria: National Botanical Institute.

CRACRAFT, J., and K. HELM-BYCHOWSKI, 1991. Parsimony and phylogenetic inference using DNA sequences: Some methodological strategies, pp. 184-220 In*: Phylogenetic Analysis of DNA Sequences (M. Miyamoto and J. Cracraft, eds)*. Oxford University Press, New York.

DANFORTH, B. N., S. SIPES, J. FANG and S. G. BRADY, 2006a. The history of early bee diversification based on five genes plus morphology. *Proceedings of the National Academy of Sciences of the USA* **103:** 15118-15123.

DANFORTH, B. N., J. FANG and S. SIPES, 2006b. Analysis of family-level relationships in bees (Hymenoptera: Apiformes) using 28S and two previously unexplored nuclear genes: CAD and RNA polymerase II. *Molecular Phylogenetics and Evolution* **39:** 358-372.

DANFORTH, B. N., and S. JI, 1998. Elongation factor-1 alpha occurs as two copies in bees: implications for phylogenetic analysis of EF-1 alpha sequences in insects. *Molecular Biology and Evolution* **15:** 225-235.

DANFORTH, B. N., H. SAUQUET and L. PACKER, 1999. Phylogeny of the bee genus *Halictus* (Hymenoptera: Halictidae) based on parsimony and likelihood analyses of nuclear EF-1 [alpha] sequence data. *Molecular Phylogenetics and Evolution* **13:** 605-618.

DANFORTH, B. N., 2002. Evolution of sociality in a primitively eusocial lineage of bees. *Proceedings of the National Academy of Sciences* **99:** 286-290.

DANFORTH, B. N., S. G. BRADY, S. D. SIPES and A. PEARSON, 2004. Single-copy nuclear genes recover Cretaceous-age divergences in bees. *Systematic Biology* **53:** 309-326.

DANFORTH, B. N., S. CARDINAL, C. PRAZ, E. A. B. ALMEIDA and D. MICHEZ, 2013. The impact of molecular data on our understanding of bee phylogeny and evolution. *Annual Review of Entomology* **58:** 57-78.

DARRIBA, D., G. L. TABOADA, R. DOALLO and D. POSADA, 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9:** 772.

DAVIES, G. B. P., and D. J. BROTHERS, 2006. Morphology of *Scrapter* (Hymenoptera: Anthophila: Colletidae), with description of three new species and taxonomic status of five Cockerell taxa. *African Invertebrates* **47:** 135-183.

DAVIES, G. B. P., C. D. EARDLEY and D. J. BROTHERS, 2005. Eight new species of *Scrapter* (Hymenoptera: Apoidea: Colletidae), with descriptions of *S. albifumus* and *S. amplispinatus* females and a major range extension of the genus. *African Invertebrates* **46:** 141-179.

DAVIES, G. B. P., 2006. Phylogenetic Systematics of *Scrapter* (Hymenoptera: Anthophila: Colletidae), Discipline of Entomology, School of Biological and Conservation Sciences. University of KwaZulu-Natal. Pitermaritzburg. South Africa Unpublished MSc dissertation.

DE QUEIROZ, A., and J. GATESY, 2006. The supermatrix approach to systematics. *Trends in Ecology & Evolution* **22:** 34-41.

DONOGHUE, M. J., J. A. DOYLE, J. GAUTHIER, A. G. KLUGE and T. ROWE, 1989. The importance of fossils in phylogeny reconstruction. *Annual Review of Ecology and Systematics* **20:** 431-460.

DRESSLER, R. L., 1982. Biology of the orchid bees (Euglossini). *Annual Review of Ecology and Systematics* **13:** 373-394.

EARDLEY, C. D., and R. URBAN, 2010. Catalogue of Afrotropical bees (Hymenoptera: Apoidea: Apiformes). *Zootaxa* **2455:** 1-548.

EARDLEY, C. D., 1996. The genus *Scrapter* Lepeletier and Serville (Hymenoptera: Colletidae). *African Entomology* **4:** 37-92.

ENGEL, M. S., 2001. A monograph of the Baltic amber bees and evolution of the Apoidea (Hymenoptera). *Bulletin of the American Museum of Natural History* **259:** 1-192.

FARRIS, J. S. 1969. A successive approximations approach to character weighting. *Systematic Zoology* **18:**374-385

FELSENSTEIN, J., 2004. PHYLIP (Phylogeny Inference Package), J. Felsenstein, Department of Genome Sciences, University of Washington, Seattle.

GATESY, J., C. MATTHEE, R. DESALLE and C. HAYASHI, 2002. Resolution of a supertree/supermatrix paradox. *Systematic Biology* **51:** 652-664.

GESS, F. W., and S. K. GESS, 1993. Effects of increasing land utilization on species representation and diversity of aculeate wasps and bees in the semi-arid areas of southern Africa. Pp 83-113 in J. LaSelle and I. D. Gauld, editors. *Hymenoptera and biodiversity*. CAB International, Wallingford, United Kingdom.

GIVNISH, T., and K. SYTSMA, 1997. Consistency, characters, and the likelihood of correct phylogenetic inference. *Molecular Phylogenetics and Evolution* **7:** 320-330.

GOLOBOFF, P., 1999. Nona Computer Program and Software, published by the author. Tücuman, Argentina. http://www.cladistics.com

GOLOBOFF, P., FARRIS, S., AND K. NIXON, 2000. TNT (Tree analysis using New Technology).

HALL, T. A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98NT. *Nucleic Acids* **41:** 95-98.

HASEGAWA, M., KISHINO, H., AND T. A.YANO, 1999. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution.* **22(1):** 160-174.

HEDGES, S. B., and L. R. MAXSON, 1996. Re: Molecules and morphology in amniote phylogeny. *Molecular Phylogenetics and Evolution* **6:** 312-314.

HOVEMANN, B., S. RICHTER, U. WALLDORF and C. CZIEPLUCH, 1998. Two genes encoded related cytoplasmic elongation factors 1alpha (EF-1alpha) in *Drosophila melanogaster* with continuous and stage specific expression. *Nucleic Acids Research* **16:** 3175-3194.

HUELSENBECK, J. P., 1991. When are fossils better than extant taxa in phylogenetic analysis? *Systematic Zoology* **40:** 458-469.

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

JEFFREYS, H., 1961. Theory of probability.3rd Edition. *Oxford Classic Texts in the Physical Sciences.* Oxford University Press. Oxford

JOHANNSMEIER, M. F., 2001. Beekeeping in South Africa. ARC-Plant Protection Research Insitute. Pretoria.

KASS, R. E., B. P. CARLIN, A. GELMAN and R. M. NEAL, 1998. Markov chain monte carlo in practice: A roundtable discussion. *The American Statistician* **52:** 93-100.

KEVAN, P. G., and T. P. PHILLIPS, 2001. The economic impacts of pollinator declines: an approach to assessing the consequences. *Conservation Ecology* (http://www.consecol.org/vol5/iss1/art8).

KIMURA, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16:** 111-120.

KLEIN, A. M., B. E. VAISSIERE, J. H. CANE, I. STEFFAN-DEWENTER and S. A. CUNNINGHAM, 2007. Importance of pollinators in changing landscapes for world crops. *Procceedings of the Royal Society of. Biology*. **274:** 303-313.

KOCH, H., 2010. Combining morphology and DNA barcoding resolves the taxonomy of western Malagasy *Liotrigona* Moure, 1961 (Hymenoptera: Apidae: Meliponini). *African Invertebrates* **51:** 413-421.

LAMBOY, W. F., 1994. The accuracy of the maximum parsimony method for phylogeny reconstruction with morphological characters. *Systematic Botany* **19:** 489-505.

LARKIN, M. A., G. BLACKSHIELDS, N. BROWN, R. CHENNA, P. A. MCGETTIGAN *et al.*, 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23:** 2947-2948.

LEISTNER, O. A., 2000. *Seed plants of Southern Africa: Families and Genera*. In *Strelitzia.* **10:** 1-775 National Botanical Institute, Pretoria

LEPELETIER de SAINT-FARGEAU, A. L. M., and A. SERVILLE, 1828. In: Latreille, P.A., ed., Encyclopèdie Mèthodique: Histoire naturelle. *Entomologie, ou histoire naturelledes Crustacès, des Arachnides et des Insects* **10**

LEWIS, P. O., 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* **50:** 913-925.

LOSEY, J. E., and M. VAUGHAN, 2006. The economic value of ecological services provided by insects. *Bioscience* **56:** 311-323.

LOW, A. B., 1998. *Vegetation of South Africa, Lesotho and Swaziland*. Department of Environmental Affairs & Tourism. Pretoria

MADDISON, W. P., and D. MADDISON, 2001. Mesquite: a modular system for evolutionary analysis. http://mesquiteproject.org/mesquite/mesquite.html.

MARDULYN, P., and S. A. CAMERON, 1999. The major opsin in bees (Insecta: Hymenoptera): a promising nuclear gene for higher level phylogenetics. *Molecular Phylogenetics and Evolution* **12:** 168-176.

MARDULYN, P., and J. B. WHITFIELD, 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.

MCGINLEY, R. J., 1981. *Systematics of the Colletidae based on Mature Larvae with Phenetic Analysis of Apoid Larvae (Hymenoptera, Apoidea)*. University of California Press. Los Angeles.

MELO, G. A. R., and R. B. GONÇALVES, 2005. Higher-level bee classifications (Hymenoptera: Apoidea: Apidae *sensu lato*). *Revista Brasileira de Zoologia* **22:** 153-159.

MICHENER, C. D., 1944. Comparative external morphology, phylogeny, and a classification of the bees (Hymenoptera). *Bulletin of the American Museum of Natural History* **82:** 151-326.

MICHENER, C. D., 1974. *The Social Behavior of the Bees: a Comparative Study*. Harvard University Press. Cambridge, Massachusetts.

MICHENER, C. D., 2000. *The Bees of the World*. First edition. The Johns Hopkins University Press, Baltimore, CA.

MICHENER, C. D., 2007. *The Bees of the World*. Second edition. The Johns Hopkins University Press, Baltimore, CA.

MICHEZ, D., C. EARDLEY, M. KUHLMANN and S. PATINY, 2007. Revision of the bee genus *Capicola* (Hymenoptera: Apoidea: Melittidae) distributed in the southwest of Africa. *European Journal of Entomology* **104:** 311-340.

MICHEZ, D., M. S. PATINY, and B. N. DANFORTH, 2009. Phylogeny of the bee family Melittidae (Hymenoptera: Anthophila) based on the combined molecular and morphological data. *Systematic Entomology* **34:** 574-597.

MILTON, S., R. YEATON, W. DEAN and J. VLOK, 1997. Succulent Karoo, pp. 131-166 in *Vegetation of southern Africa*. Cambridge University Press, Cambridge, UK.

MIYAMOTO, M. M., 1985. Consensus cladograms and general classifications. *Cladistics* **1:** 186-189.

MIYAMOTO, M. M., and J. CRACRAFT, 1991. Phylogenetic inference, DNA sequence analysis and the future of molecular systematics., In*: Miyamoto, M. M., Cracraft, J. (Edition), Phylogenetic Analysis of DNA Sequences.* Oxford University Press, New York.

MIYAMOTO, M. M., and W. M. FITCH, 1995. Testing species phylogenetics and phylogenetic methods with congruence. *Systematic Biology* **44:** 64-76.

NEFF, J. L., and B. B. SIMPSON, 1981. Oil-collecting structures in the Anthophoridae (Hymenoptera): morphology, function, and use in systematics. *Journal of the Kansas Entomological Society* **54:** 95-123.

NOVACEK, J. M., 1992. Fossils, topologies, missing data, and the higher level phylogeny of eutherian mammals. *Systematic Biology* **41:** 58-73.

NYLANDER, J., 2004. MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

POSADA, D., 2003. Using MODELTEST and PAUP* to select a model of nucleotide substitution. Current Protocols in Bioinformatics**:** DOI: 10.1002/0471250953.bi0605s00

RAIMONDO, D., L. V. STADEN, W. FODEN, J. VICTOR, and N. HELME. 2009. *Red list of South African plants 2009*. South African National Biodiversity Institute.

RAMBAUT, A., 2009. FigTree, version 1.3.1. Institute of Evolutionary Biology, University of Edinburgh. http://tree.bio.ed.ac.uk/

RAMBAUT, A., and A. J. DRUMMOND, 2009. Tracer, version 1.5. MCMC Trace Analysis Package, Institute of Evolutionary Biology, University of Edinburgh. http://www.beast.bio.ed.ac.uk/

ROUDER, J. N., and R. D. MOREY, 2012. Default Bayes factors for model selection in regression. *Multivariate Behavioral Research* **47:** 877-903.

SANDERSON, M. J., A. PURVIS and C. HENZE, 1998. Phylogenetic supertrees: Assembling the tree of life. *Trends in Ecology & Evolution* **13:** 105-109.

SCHUBERT, M., L. Z. HOLLAND, N. D. HOLLAND and D. K. JACOBS, 2000. A phylogenetic tree of the Wnt genes based on all available full-length sequences, including five from the cephalochordate amphioxus. *Molecular Biology and Evolution* **17:** 1896-1903.

SIDOW, A., 1992. Diversification of the Wnt gene family on the ancestral lineage of vertebrates. *Proceedings of the National Academy of Sciences of the USA* **89:** 5098-5102.

SIMONSEN, T. J., N. WAHLBERG, A. V. BROWER and R. DE JONG, 2006. Morphology, molecules and fritillaries: approaching a stable phylogeny for Argynnini (Lepidoptera: Nymphalidae). *Insect Systematics & Evolution* **37:** 405-418.

SMIT, H., T. ROBINSON, J. WATSON and B. J. VAN VUUREN, 2008. A new species of elephant-shrew (Afrotheria: Macroscelidea: Elephantulus) from South Africa. *Journal of Mammalogy* **89:** 1257-1268.

SOLTIS, P. S., and D. E. SOLTIS, 2004. The origin and diversification of angiosperms. *American Journal of Botany* **91:** 1614-1626.

SRINGER, M. S., R. W. DEBRY, C. DOUADY, H. M. AMRINE, O. MADSEN, W. W. DEJONG and M. J. STANHOPE. 2001. Mitochondrial Versus Nuclear Gene Sequences in Deep-Level Mammalian Phylogeny Reconstruction. *Molecular Biology and Evolution* **18(2):** 132-143.

TAMURA, K., and M. NEI, 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* **10:** 512-526.

TAMURA, K., D. PETERSON, G. STECHER, M. NEI and S. KUMAR, 2011. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution.* http://www.kumarlab.net/publications.

THORP, R. W., 1979. Structural, behavioral, and physiological adaptations of bees (Apoidea) for collecting pollen. *Annals of the Missouri Botanical Garden* **66:** 788–812.

THORP, R. W., 2000. The collection of pollen by bees. *Plant Systematics and Evolution* **222:** 211–223.

VAN WYK, A. E., and G. F. SMITH, 2001. Regions of Floristic Endemism in Southern Africa: a review with emphasis on succulents. Hatfield, South Africa: Umdaus Press viii, 199p. col. illus. ISBN 1919766189.

WARD, P. S., and S. G. BRADY, 2003. Phylogeny and biogeography of the ant subfamily Myrmeciinae (Hymenoptera: Formicidae). *Invertebrate Systematics* **17:** 361-386.

WARREN, A. D., J. R. OGAWA and A. V. BROWER, 2009. Revised classification of the family Hesperiidae (Lepidoptera: Hesperioidea) based on combined molecular and morphological data. *Systematic Entomology* **34:** 467-523.

WCISLO, W. T., and J. H. CANE, 1996. Floral resource utilization by solitary bees (Hymenoptera: Apoidea) and exploitation of their stored foods by natural enemies. *Annual Review of Entomology*. **41:** 257–286.

WHITE, F., 1983. The vegetation of Africa: A Descriptive Memoir to accompany the UNESCO/AETFAT/UNSO Vegetation Map of Africa (Natural Resources Research 20). UN Educational, Scientific and Cultural Organization, Paris.

WIEGMANN, B. M., C. MITTER, J. C. REGIER, T. P. FRIEDLANDER, D. M. WAGNER *et al.*, 2000. Nuclear genes resolve Mesozoic-aged divergences in the insect order Lepidoptera. *Molecular Phylogenetics and Evolution* **15:** 242-259.

WIENS, J. J., 2003. Missing data, incomplete taxa, and phylogenetic accuracy. *Systematic Biology* **52:** 528-538.

WIENS, J. J., and M. C. MORRILL, 2011. Missing data in phylogenetic analysis: reconciling results from simulations and empirical data. *Systematic Biology* **60:** 719-731.

WIENS, J. J., and T. W. REEDER, 1995. Combining data sets with different numbers of taxa for phylogenetic analysis. *Systematic Biology* **44:** 548-558.

WILKINSON, M., 1995. Coping with abundant missing entries in phylogenetic inference using parsimony. *Systematic Biology* **44:** 501-514.

ZWICKL, D., 2006. GARLI, vers. 0.951. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. PhD. dissertation, University of Texas, Austin, Texas, USA.