BRAIN-be
Belgian Research Action through Interdisciplinary Networks

PIONEER PROJECTS

SYRPHIDAE PLANT INTERACTIONS INTRODUCING NEXT GENERATION SEQUENCING TECHNOLOGIES

CONTRACT - BR/314/PI/SYRPINTINE

FINAL REPORT

15/02/2018

Promotors
Kurt Jordaens (Royal Museum for Central Africa, Leuvensesteenweg 13, B-3080 Tervuren, Belgium)
Steven Janssens (Botanic Garden Meise, Domein van Bouchout, Nieuwelaan 38, 1860 Meise, Belgium)

Authors
Kurt Jordaens (Royal Museum for Central Africa)
Steven Janssens (Botanic Garden Meise)
Yannick De Smet (Royal Museum for Central Africa)

Published in 2018 by the Belgian Science Policy
Avenue Louise 231
Louizalaan 231
B-1050 Brussels
Belgium
Tel: +32 (0)2 238 34 11 – Fax: +32 (0)2 230 59 12
http://www.belspo.be

Contact person: Georges JAMART
+32 (0)2 238 36 90

Jordaens, K., Janssens, S. & De Smet, Y. - ***Syrphidae Plant Interactions Introducing Next Generation Sequencing Techniques*** - Final Report. Brussels : Belgian Science Policy 2018 – 50 p. (BRAIN-be - Belgian Research Action through Interdisciplinary Networks)

TABLE OF CONTENTS

SUMMARY

Context

One of the emerging fields in the management and conservation of biodiversity is the study of plant-pollinator networks. Surprisingly, most of the efforts to understand plant-pollinator networks is dedicated to well-known insect groups such as Hymenoptera (bees and wasps), Lepidoptera (butterflies) and Coleoptera (Beetles), whereas the role of Diptera (two-winged insects, i.e. flies and mosquitoes) in these networks is strongly neglected, and this certainly holds for the Afrotropical Region. As a result there is no baseline to describe the current status of the role of Diptera in plant-pollinator networks. Pinpointing the relative role of each of the Diptera groups, or species, in plant pollination is very challenging and strongly hampered by a remarkable taxonomic impediment. Afrotropical hoverflies are one of the important pollinator groups in the Afrotropics but their taxonomy is very poor so that it remains impossible to estimate the relative contribution of each of the species in plant-pollinator networks. The unambiguous identification of pollinators is a prerequisite in the ecological study of plant-pollinator networks. Moreover, at the same moment plants (and/or their pollen) should be identified unambiguously and, as for the pollinators, this requires a stable taxonomy. Also this aspect of plant-pollinator networks is still in its infancy in the Afrotropical Region.

Objectives

In order to set an initial baseline for to role of hoverflies (Syrphidae) in plant-pollinator networks in the Afrotropics, we simultaneously tried to improve pollinator taxonomy and identification, and created a first palynological reference database for pollinating plants, both through the use of novel, molecular and state-of-the-art next-generation sequecing (NGS) techniques. The target pollinator family was the hover fly genus *Eristalinus* and the region where the research was conducted was southern Benin.

More specifically, we aimed:

- to collect, identify and establish a large reference collection for the common Afrotropical hover fly genus *Eristalinus*.
- to explore the Targeted Amplicon Sequencing technique to simultaneously sequence six DNA regions that can be used to delimit species boundaries in the genus.
- to create a reference palynological database of flowering plants in S Benin.
- to explore the Targeted Amplicon Sequencing technique to simultaneously sequence four DNA regions that can be used to identify (Afrotropical) plant species and to explore the technique to identify plant pollen.
- to link the results obtained from the taxonomic study on the hover flies with the results obtained from the molecular and palynological studies of the flowering plants to establish a baseline for plant-pollinator studies in S Benin.

Conclusions

1) We obtained a reference collection of over 800 *Eristalinus* specimens of about 30 species, not only for S Benin, but for the entire Aftrotropics.
2) The TAS-protocol proved not successful to simultaneously sequence six genes in the genus. Rather, we successfully used classical Sanger Sequencing to delimit species in the genus.
3) Alternatively, we sequenced the entire mtDNA genome of five *Eristalinus* species.
4) These five mtDNA mitogenomes will be used as a backbone for an ongoing large-scale High Throughput Sequencing study where full mtDNA genomes will be obtained from 120 species of hover flies of the Afrotropics (of the genera *Eristalinus*, *Syritta*, *Phytomia*, *Xylota*, *Mesembrius*).
5) We established a first DNA barcode reference database for the flowering plants in S Benin.
6) We established a first palynological reference database for the flowering plants in S Benin.
7) The TAS protocol proved very successful to identify flowering plant species from S Benin, yet, we need to improve the technique so that it can be used to identify pollen collected from hover flies.
8) The protocol to collect pollen from hover flies in the Afrotropics requires improvements before the relative contribution of each species in plant-pollinator networks can be estimated.

Keywords

Afrotropics, Benin, DNA barcoding, flower flies, flowering plants, hover flies, mtDNA, palynology, plant-pollinator networks, pollen (research), species delimitation, species identification, target-amplicon sequencing, whole mitochondrial DNA sequencing.

SAMENVATTING

Context

Eén van de opkomende onderzoeksdomeinen in de bescherming en het beheer van onze biodiversiteit is het onderzoek naar plant-bestuivers netwerken. Opmerkelijk genoeg gaat de meeste van onze aandacht naar goed gekende en veel bestudeerde insectengroepen zoals Hymenoptera (bijen en wespen), Lepidoptera (vlinders) en Coleoptera (kevers), daar waar de rol van Diptera (tweevleugelige insecten, i.e. vliegen en muggen) in deze netwerken, zeker in de Afrotropische regio, nauwelijks geapprecieerd wordt. Een gevolg hiervan is dat de rol en het belang van Diptera in plant-bestuiver netwerken niet voldoende ingeschat kan worden. Het inschatten van deze rol is ook niet eenvoudig aangezien het inschatten van het relatieve belang van elke vliegensoort een correcte identificatie van de insecten veronderstelt en dit laatste wordt sterk bemoeilijkt door een gebrek aan taxonomische kennis over de Diptera van de Afrotropische regio. Afrotropische zweefvliegen zijn een van de belangrijkste groepen van bestuivers in de Afrotropische regio maar hun taxonomie is onvoldoende gekend en daarom is het momenteel onmogelijk om hun bijdrage en rol in plant-bestuivers netwerken te kunnen inschatten. Immers, een correcte identificatie van bestuivers is noodzakelijk om de ecologische relaties tussen bestuivers en bloeiende planten in kaart te kunnen brengen. Bovendien moeten ook de plantensoorten juist geïdentificeerd kunnen worden en moet het pollen dat de bestuivers met zich meedragen en wat voor de bestuiving van de bloemen zorgt, gelinkt kunnen worden aan de juiste plantensoort. Ook dit laatste aspect staat nog in haar kinderschoenen in de Afrotropische regio.

Doelstellingen

Ten einde een basislijn te kunnen opstellen voor de rol van zweefvliegen (Syrphidae) in plant-bestuivers netwerken in the Afrotropische regio, hebben we tegelijkertijd geprobeerd de taxonomie van de bestuivers op te helderen en hun identificatie te vergemakkelijken, als om een referentie-gegevensbank van bloeiende platen en hun pollen voor Zuid-Benin op te stellen, dit door gebruik te maken van recent ontwikkelde hoog-geavanceerde DNA sequentietechnieken (zogenaamde next-generation DNA sequentie technieken). Onze doelgroep binnen de Afrotropische zweefvliegen was het genus *Eristalinus* en het onderzoek spitste zich toe op de region van Zuid-Benin.

Meer specifiek waren onze doelstellingen:

- het verzamelen, identificeren en opstellen van een grote referentie-gegevensbank van de Afrotropische soorten van het zweefvliegengenus *Eristalinus*.
- het evalueren van de Targeted Amplicon DNA sequeneringstechniek voor het simultaan sequeneren van zes specifieke regio's van het genoom van deze zweefvliegen om de verschillende soorten van dit genus te omlijnen en om hun identificatie te vereenvoudigen.
- het aanmaken van een referentie palynologische gegevensbank van de bloeiende planten van Zuid-Benin.

- het evalueren van de Targeted Amplicon DNA sequeneringstechniek voor het simultaan sequeneren van vier specifieke regio's van het genoom van deze plantensoorten om deze te kunnen identificeren en om deze techniek toe te passen op de identificatie van pollen.
- de resultaten van de taxonomische studies op de zweefvliegen te integreren met de resultaten bekomen uit de palynologische en moleculaire studies op de pollen en planten om op deze manier een eerste basislijn voor plant-bestuivers netwerken op te stellen voor Zuid-Benin.

Besluiten

1) We hebben momenteel een referentie-verzameling van meer dan 800 *Eristalinus* individuen van ongeveer 30 soorten, niet enkel voor Zuid-Benin, maar voor de gehele Afrotropische regio.
2) Het TAS-protocol bleek niet succesvol in het simultaan sequeneren van zes genen binnen het genus. Daarom gebruikten we klassieke Sanger DNA sequenering om de soorten binnen het genus Eristalinus te omlijnen en om hun identificatie te vergemakkelijken.
3) Als alternatief hebben we ook vijf volledig mitochondriale genomen gesequeneerd.
4) Deze vijf mitochondriale genomen doen dienst als ruggengraat voor een lopende, grootschalige DNA sequentie-analyse van volledige mitochondriale genomen van 120 zweefvliegensoorten uit de Afrotropische region (uit de genera *Eristalinus*, *Syritta*, *Phytomia*, *Xylota*, *Mesembrius*).
5) We maakten een eerste referentie-gegevensbank van DNA barcodes voor de bloeiende planten van Zuid-Benin.
6) We maakten een eerste referentie palynologische gegevensbank voor de bloeiende planten van Zuid-Benin.
7) Het TAS protocol bleek zeer succesvol voor de identificatie van de bloeiende planten van Zuid-Benin maar de protocols voor het gebruik van deze techniek voor het identificeren van pollenkorrels moeten nog beter aangepast en ontwikkeld worden.
8) Het protocol voor het verzamelen van pollen van insecten uit tropische regio's moet ook verder op punt gesteld worde om zodoende de relatieve bijdrage van de verschillende zweefvliegen soorten in plant-bestuivers netwerken in Zuid-Benin te kunnen inschatten.

Trefwoorden

Afrotropische regio, Benin, bloeiende planten, DNA barcoding, DNA sequentie-analyse, mtDNA, mitochondriaal DNA, mitochondriaal DNA sequentie-analyse, mtDNA, palynologie, plant-pollinator networks, pollen(onderzoek), species delimitation, species identification, target-amplicon DNA sequenering, soort-identificatie, soort-omlijning, zweefvliegen

1. INTRODUCTION

The pollination of flowering plants by animals is critical to agriculture and thus vitally important to humanity. Seventy-five percent of all crop species depend on animal pollinators to produce fruits or seeds (Klein et al., 2007). Conserving these essential and valuable pollination services is crucial to achieve food supply security. Nowadays, pollination systems are under increasing threat from anthropogenic sources, including fragmentation of habitat, changes in land use, modern agricultural practices, and use of chemicals such as pesticides and herbicides. For instance, honeybee numbers are dramatically decreasing worldwide (e.g. van Engelsdorp et al., 2009; Cameron et al., 2011) and address the importance of healthy pollination systems. Most attention has been given to managed honeybee populations (Allsopp et al., 2008; Kremen, 2008).

However, introduced honeybees may harm native bees and other pollinators (Kremen et al., 2002) urging the further development of native animals as crop pollinators. Indeed, the "pollination crisis" resulting from a decline of honeybees (and native bees) may be compensated not only by changes in habitat use and agricultural practices but also by cultivation of a diversity of crop pollinators (Winfree et al., 2007; Klein et al., 2009). This requires increased efforts to study basic aspects of plant-pollinator interactions if optimal management decisions are to be made for conservation of these interactions in natural agricultural ecosystems. After all, biotic interactions play an important role on the organization of pollinating systems, and persistence of biodiversity in general.

One of the emerging fields in the management and conservation of biodiversity is the study of plant-pollinator networks. Surprisingly, most of the efforts to understand plant-pollinator networks is dedicated to well-known insect groups such as Hymenoptera (bees and wasps), Lepidoptera (butterflies) and Coleoptera (Beetles), whereas the role of Diptera (two-winged insects, i.e. flies and mosquitoes) in these networks is strongly neglected, and this certainly holds for the Afrotropical Region. As a result there is no baseline to describe the current status of the role of Diptera in plant-pollinator networks. However, plant-pollinator networks are very complex, not only because of interactions of a variety of pollinators with a variety of flowering plant species, but also because of interactions among different taxa of pollinators such as Hymenoptera (bees and wasps), Lepidoptera (butterflies), Coleoptera (Beetles), and Diptera, and because of interactions among species and specimens of the same taxon. If we want to fully understand how plant-pollinator networks evolve, and change under changing environmental conditions, we need to know the role of the Diptera in these networks. Only 1/2 to 2/3 of the Diptera species of the Afrotropical Region have currently been described. Pinpointing the relative role of each of the Diptera groups, or species, in plant pollination is thus very challenging and strongly hampered by this taxonomic impediment. Moreover, much of the taxonomic information, including morphological identification keys, is scattered throughout the literature, sometimes difficult to access, and often incomplete. A case in point are the Afrotropical hoverflies (Figure 1) for which the taxonomy is very poor. Besides, a correct identification of pollinators is necessary to estimate the relative contribution of each of the species in plant-pollinator networks and to unravel these networks. Hence, the unambiguous identification of pollinators is a prerequisite in the ecological study of plant-pollinator networks.

Moreover, at the same moment plants (and/or their pollen) should be identified unambiguously and, as for the pollinators, this requires a stable taxonomy.Also this aspect of plant-pollinator networks is still in its infancy in the Afrottropical Region.

In order to set an initial baseline for to role of hoverflies (Syrphidae) in plant-pollinator networks in the Afrotropics, we simultaneously tried to improve pollinator taxonomy and identification, and created a first palynological reference database for pollinating plants, both through the use of novel, molecular and state-of-the-art next-generation sequecing (NGS) techniques. Before extending on the molecular techniques we first give more details on the study system, viz. Afrotropical hover flies and flowering plants.

The pollinators: Afrotropical hover flies

Detailed knowledge about the life history and ecology of pollinators is essential to study interactions with their food plants. There are over 5000 hover fly (Diptera, Syrphidae) species worldwide, of which the majority feed on pollen and nectar. Thus, hover flies (also known as flowerflies) represent one of the major plant pollinator insect guilds.

Whereas the taxonomy of this insect family is relatively well-known for several biogeographic regions, that of the Afrotropical region remains largely unexplored. Nevertheless, with its 500 and more morphospecies (Dirickx, 1998), hover flies are a significant biodiversity component of the African entomofauna. Unfortunately, little is known of their feeding and pollination biology and a major reason for this is that species are ill-defined resulting in a very unstable and unsatisfactory taxonomy. A reliable species identification and delimitation is thus fundamental and pivotal to study plant-pollinator interactions, and to numerous other biological disciplines such as evolution, macro-ecology, and conservation.

African Syrphidae are traditionally identified on the basis of morphological differences. Especially the presence or absence of hairs on different body parts, color differences of various body components, and difference in wing venation are used to identify species. Yet, differences between related species can be very subtle and many species cannot be identified unambiguously on morphological grounds. This also holds for one of the most common genera of African Syrphidae, viz. *Eristalinus*, that was selected for this study (Figure 1). Species of this genus are commonly found on flowering wild and agricultural plants.

Eristaline (tribe Eristalini) flower flies are common flower visitors whose larvae are aquatic and characterized by a long anal breathing tube (hence, the common name "rat-tailed" maggots). The adults are useful pollinators and the maggots filter and purify water. Species of this tribe are characterized by: 1) a pilose postpronotum, 2) a strongly sinuate vein r 4+5, 3) a closed and petiolate cell R1, 4) a metafemur with a basolateral patch of black setulae, 5) an anepimeron with a pilose triangular portion, 6) a postalar pile tuft, 7) a wing hyaline, and 8) a scutellum without an apical marginal sulcus. The best field character is that the eyes are always with distinctive markings, either

maculae or fasciae or a combination of both. Unfortunately, in some preserved specimens the eyes turn completely dark but the original coloration can usually be temporarily restored by wetting the eyes with alcohol or ethyl acetate.

The genus *Eristalinus* is one of the most speciose group in the Afrotropics with more than 55 currently recognized morphospecies divided into five subgenera (Thompson 2003; De Smet et al., in prep.) (Figure 1). Unfortunately, the only catalogue to Afrotropical flies (Smith & Vockeroth 1980) did not recognize these subgroups, and the later catalogue to the flower flies of the region (Dirickx 1998) only recognized two of the four subgenera. Finally, Whittington (2003) recommended the return to a single genus until a revision was done.

Identification of the species of this group is difficult especially as the last key to these species was by Curran (1939, as *Lathyrophtalmus* and *Eristalodes*), which in turn was based on the only other key to species (Bezzi 1915, covering 18 species). Curran (1939) only keyed 30 out of the then 48 known species.

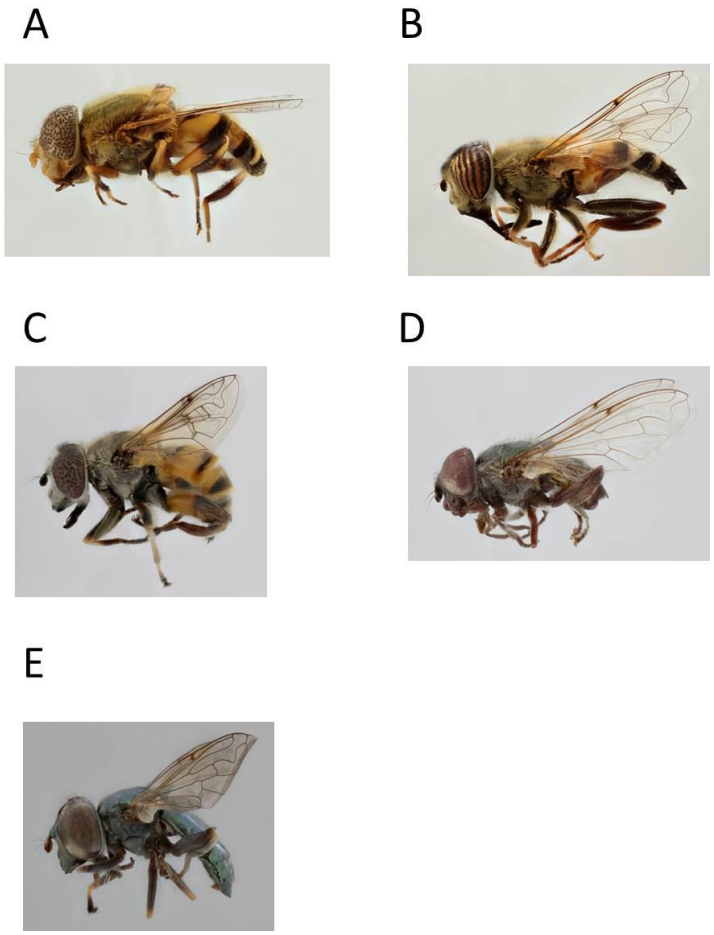

Figure 1: Representatives of the five subgenera of the Aftotropical species of the hover fly genus *Eristalinus*: A. subgenus *Eristalinus* s.s.; B. subgenus *Eristalodes* ; C. subgenus *Merodonoides*; D. subgenus *Oreristalis*; E. subgenus *Helophilina*.

The plants: Central African flowering plants of agricultural importance

Benin is predominantly an agricultural country. About 55% of the economically active population was engaged in the agricultural sector in 2000, which accounted for 38% of GDP that year. Small, independent farmers produce 90% of the agricultural products (Figure 2) and as such agricultural activities are interspersed with (semi) natural habitat. The conservation of pollinator services by insects is thus of major importance for seed and food production in this area. Unfortunately, the need for active conservation of pollination interactions is currently not being appreciated.

So far, basic plant-pollinator interactions have been studied through 1) the observation of the flower visit behaviour of insects and 2) palynology, which is the study of plant pollen. Pollen recovered from the gut or intestinal tract of insects may be used to understand the diets of insects because the shape and size of pollen is often species-specific (Dessein et al., 2005; Janssens et al., 2012). Palynological studies, however, are time-consuming and can only be performed by specialists. Moreover, different plant species may show very similar pollen morphology thus hampering the correct identification of food plants impeding a correct interpretation of plant-pollinator relationships.

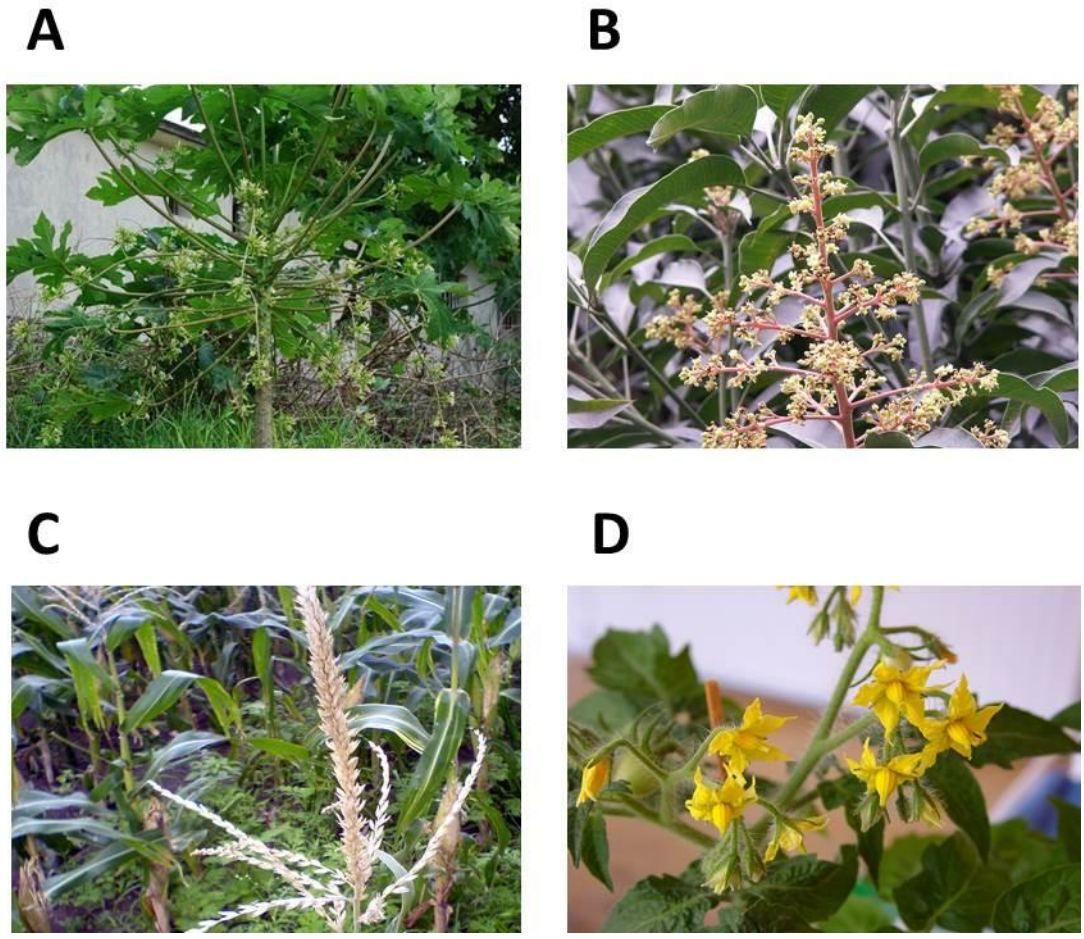

Figure 2: Flowers of four common agricultural crops in Benin: A. Papaya (*Carica papaya*); B. mango (*Mangifera indica*); C. maize (*Zea mays*); D. tomato (*Solanum* sp.).

2. METHODOLOGY AND RESULTS

Collection of Syrphidae

We have collected adult hover flies during several sampling campaigns over the last years: Benin (2013, 2015, 2016), Togo (2016), Malawi (2016), Kenya (2016), South Africa (2017) and La Réunion (2015). Hover flies were collected using insect nets (Figure 3A, B and C) and with Malaise traps (Figure 3D). In addition, we received *Eristalinus* specimens from several colleagues who allowed us to use the specimens in our molecular analysis. As such, we received material from Gunilla Ståhls, Ximo Mengual and Axel Ssymank (South Africa), Georg Goergen (Benin, Togo, Ghana, Seychelles), Martin Hauser (Madagascar), Laban Njoroge and Robert Copeland (Kenya), and Burgert Muller and Ashley H. Kirk-Spriggs (South Africa, Malawi, La Réunion, Benin, Togo). This resulted in a collection of more than 800 *Eristalinus* specimens that were identified morphologically. After morphological examination, we selected representative specimens for each morphospecies for molecular analysis.

Collection of pollen and leaf material

Most of the sampling for the pollen analysis was carried out on the IITA (International Institute of Tropical Agriculture) compound in Cotonou (Benin) where there is a large variety of different (native and non-native) cultivated crops varying from mango (*Mangifera indica*), plantain (*Musa ABB*), avocado (*Persea americana*), papaya (*Carica papaya*) and maize (*Zea mays*).

During the field work in Togo and Benin (2016) (Figure 3A, B and C), we collected approximately 50 *Eristalinus* specimens on flowering plant species. *Eristalinus* specimens were individually stored in 2D-barcoded tubes that contained CTAB buffer. Each time, pollen was removed from the specimens using a wet tissue and the tissue was rinsed in CTAB-buffer; the rinsed product was also stored in individual tubes. From each of the flowering plant species where *Eristalinus* specimens were collected from, a small piece of leaf tissue was kept in a coffee filter and stored in a plastic bag to which silica crystals were added to prevent moulding of the leaf tissue. Upon arrival, leaf tissues, pollen and syrphids were kept at the Botanical Garden Meise for molecular examination and palynology. Flowering plants were identified either on field or via photographs and herbarium material. Field identifications were verified in Botanic Garden Meise by using the large herbarium collection of African plants as reference collection.

Figure 3: Photographs of field work in Benin and Togo (2016): A. Yannick De Smet and Kurt Jordaens collecting hover flies from a flowering *Terminalia* tree; B. Yannick De Smet sampling flowering plant species; C. Kurt Jordaens and Yannick De Smet collecting hover flies in the botanical garden of Porto Novo ; D. Malaise trap over a small river bed in Togo.

DNA sequencing technologies

Modern DNA sequencing techniques are increasingly used to improve the taxonomy of a wide variety of insect taxa. Yet, statistical species delimitation usually relies on singular data (i.e. DNA sequences of a single gene or gene fragment), primarily genetic, for detecting putative species and individual assignment to putative species. Given the variety of speciation mechanisms, singular data may not adequately represent the genetic, morphological and ecological diversity relevant to species delimitation. Therefore, we need a methodological framework combining multivariate and clustering techniques that uses molecular, morphological and ecological data to detect and assign individuals to putative species. Below, we give more information on the NGS-techniques that we considered during the SYRPINTINE project (Anchored phylogeny, RAD-Seq, TAS).

Anchored phylogeny

Currently, a consortium studies the phylogenetic relationships of all syrphid genera worldwide using an NGS-technique called "Anchored-phylogeny". Although this technique seems highly suitable to reveal both deep and shallow phylogenetic relationships among syrphid taxa (it has resulted in one publication on the phylogeny of Syrphidae in which Kurt Jordaens is co-author), the technique is very expensive, time-consuming, and requires many man-hours to be completed. Hence, we considered the technique to be out-of-scope for the SYRPINTINE project.

DNA barcoding using Target-Amplicon Sequencing (TAS)

Recent advances in molecular DNA-sequencing techniques allow the identification of animals and plants using short DNA fragment, called DNA-barcodes (Hebert et al., 2003). DNA-barcoding uses Sanger sequencing of single gene fragments of individual specimens to generate large-scale DNA barcode libraries. These libraries are then used to identify unknowns. However, direct Sanger sequencing of DNA barcode amplicons, as practiced in most DNA barcoding procedures, is hampered by the need for relatively high-target amplicon yield, co-amplification of nuclear mitochondrial pseudogenes, and instances of intra-individual variability (heteroplasmy). Any of these situations can lead to failed Sanger sequencing attempts or ambiguity of the generated DNA barcodes.

Targeted sequencing of specific loci using NGS platforms may provide a more efficient means of generating data sets that are devote of these shortcomings. Barcode loci can be enriched prior to NGS using well-established PCR techniques or newly developed hybridization techniques (e.g. Briggs et al., 2009; Gnirke et al., 2009; Maricic et al., 2010) and then pooled for high-throughput sequencing. These targeted approaches (target amplicon sequencing or TAS) increase the sequence coverage for any individual locus and reduce the probability of missing data, which may have negative impact on data analyses (Lemmon et al., 2009). The potential for a targeted NGS approach has been further advanced by the development of indexing strategies that permit pooling and subsequent parallel tagged sequencing of multiple individual samples within a single NGS run (Binladen et al., 2007; Bybee et al., 2011). As such, TAS may be a cost-efficient, rapid, and highly reproducible identification technique for pollen collected from the stomach or other body parts of insect species.

TAS uses a two-step PCR process that allows to amplify a targeted gene region (amplicon) using traditional PCR, followed by an additional PCR that attaches a known 10 bp tag, or MID, to identify amplicons from different samples (hoverfly species or taxa, in our case). This also allows to simultaneously amplify and sequence different genes.

Currently, there is an ongoing study to unravel the phylogeny of all syrphid genera using six gene fragments, viz. AATS, 18S, 28S, CAD1,CAD4/5, COI-HEB, COI-PJ, and Period. During the start of

the SYRPINTINE project we invited to join this international study. Because the six-gene phylogenetic approach felt within the objectives of the SYRPINTINE project we focused on the TAS-technique, rather than on the RAD-seq approach that was originally proposed in the project.

However, the amplification and DNA Sanger sequencing of AATS, CAD1, CAD4/5 and Period proved very difficult and results were unsatisfactory. We therefore expected a low sequencing success for these genes using TAS. Nevertheless, we outsourced the simultaneous sequencing of these six genes to LGC Genomics (Germany) but their technical team confirmed that it was nearly impossible to sequence the six genes simultaneously using TAS (or any other related NGS-technique). We therefore abandoned the use of TAS and looked for an alternative NGS) technique (whole mitochondrial DNA sequencing) that would prove useful for studying the phylogeny of *Eristalinus* (see below). However, because whole mitochondrial sequencing DNA was entirely new for our research groups, we used muti-gene classical Sanger-Sequencing approach to be sure to have a sufficiently resolved taxonomy of the hover flies and to be able to link identifications of hover flies to correctly identified pollen and flowering plant species.

For the identification of the plant species, however, the TAS-protocol proved highly rewarding. Before we go into detail on the TAS-protocol for the identification of flowering plants species, we first explain the results and conclusions of the taxonomic work on the hover flies of the genus *Eristalinus*.

Sanger-Sequencing: Species delimitation in the hover fly genus *Eristalinus*

For the 800 *Eristalinus* specimens at our disposal, we selected 150 specimens for the molecular species delimitation study. These specimens represented 23 *Eristalinus* morphospecies, of which 19 corresponded to described putative species. Of the remaining four morphospecies, two (*Eristalinus* sp. nov. 1, *Eristalinus* sp. nov. 2) represented undescribed morphospecies, exhibiting combinations of diagnostic characters not found in any of the studied type specimens or original descriptions. The remaining two are represented by a single individual (impeding morphological identification and molecular algorithmic species delimitation).

Total genomic DNA was extracted using the DNeasy® Blood & Tissue kit (Qiagen Inc., Hilden, Germany), from a single front leg. Specimens were either stored in ethanol or pinned, and either collected for this study or received on loan from various collections. A total of 149 specimens were selected for this study, and identified by external morphology using the identification keys provided in Bezzi (1915) and Curran (1939) or based on the original species descriptions and comparison to type specimens. We included 3-5 specimens per putative taxon, in order to reflect both inter- and intraspecific variation. For each specimen, three mitochondrial (*cyt*b, COI, 12S) and two nuclear (18S, 28S) regions were sequenced. Amplification of COI utilized the primers HCO and LCO (Folmer et al., 1994), while PCR amplification of the remaining regions used M13-labeled primers. Sequencing was outsourced to Macrogen Europe, using the HCO and LCO primers for COI, but

M13f and M13r for all other regions. Sequences of the forward and reverse strands were assembled and edited in PhyDE (Müller et al., 2008), and multiple sequence alignments were generated with MUSCLE (Edgar, 2004).

Phylogenetic reconstruction

Gene trees were estimated in MrBayes, using four independent runs of eight chains each, for 10 million generations. Substitution models were selected by jModeltest (Darriba et al., 2012) for each region or, when not available, the next more complicated model implemented in MrBayes. Convergence of runs was assessed using the standard deviation of split frequencies, in addition to monitoring the ESS (effective sample size) values and distribution of log-likelihoods in Tracer v1.6 (Rambaut & Drummond, 2013). The initial 25% of trees was discarded as burn-in, while the remaining subset was used to construct a 50% majority-rule consensus tree.

The *BEAST algorithm implemented in BEAST v1.8.2 (Heled & Drummond, 2010) was used to estimate species trees under the coalescent, using all five sequenced regions or only COI + cyt*b*, since these markers showed the highest variability. Both analyses consisted of five independent BEAST runs, for 30 million generations each. For each dataset, log- and tree-files were combined respectively, across all five runs using LogCombiner v1.8.2 (Heled & Drummond, 2010). The resulting combined log files were evaluated in Tracer v1.6 to assess ESS and log-likelihood values. Before combining tree files across runs, the first 25% was discarded as burn-in. With the remaining sampled trees, a maximum clade credibility tree was generated using TreeAnnotator v1.8.2 (Heled & Drummond, 2010). All phylogenetic analyses in MrBayes and BEAST were run on the CIPRES Science Gateway computing cluster (Miller et al., 2010).

Four species delimitation programs were used to infer species boundaries and the total number of putative species in the specimen pool.

1. Generalized model Yule Coalescent

To streamline terminology within this paper, entities delimited by sequence-based algorithms will be referred to as “clusters”, instead of the specific terminology used in each method. Only the *cyt* b and COI datasets were used for the Generalized model Yule Coalescent (GMYC) algorithm, since they provided the highest interspecific divergece. The BEAST v1.8.2 package (Heled & Drummond, 2010) was used to generate ultrametric trees. Two separate analyses were run for each marker, using different tree priors, as these can influence the reconstruction of the phylogenetic trees (Ceccarelli et al., 2012), and thus the species delimitation hypotheses based on them. The different tree priors were: the coalescent tree prior (resulting trees: $UT_{COI, C}$ and $UT_{cytb, C}$), recommended for population level data, and the Yule prior (resulting trees: $UT_{COI, Y}$ and $UT_{cytb, Y}$) , generally recommended for species-level data (Huw et al., 2016). Each of these analyses was run for 30 million generations, with the uncorrelated lognormal relaxed clock model, and the nucleotide substitution model selected in jModeltest. Convergence and sample sizes were evaluated in Tracer v1.6, and for each analyses the first 25% of sampled trees were discarded before producing a maximal clade credibility tree in TreeAnnotator v1.8.2. All ultrametric trees

($UT_{COI, Y}$, $UT_{COI, C}$, $UT_{cytb, Y}$ and $UT_{cytb, C}$) served as input for the species discovery algorithm GMYC which is implemented in the SPLITS package (http://r-forge.r-project.org/projects/splits/) for R (R Core Development Team, 2013). Each tree was analyzed using both the single and multiple threshold options, generating four species hypotheses for each of the two markers.

2. Poisson tree Processes

Poisson tree processes (PTP, Zhang et al., 2013) analyses took Bayesian phylogenetic hypotheses for markers COI and *cyt* b as input. The remaining markers were not used in this analyses because they showed low resolution at the species level. Each of these analyses was run on the bPTP web server (http://species.h-its.org/), which implements a Bayesian version of PTP, adding posterior probabilities to putative species. Each run consisted of 500.000 generations, of which the first 10% were discarded in calculating the posterior speciation probabilities. Convergence was assessed using the generated log-likelihood plot.

3. Automatic barcode gap discovery

Automatic barcode gap discovery (ABGD) takes distance matrices or multiple sequence alignments as input, to recursively split sequences into clusters based on sequence divergence (Puillandre et al., 2011). Here, the COI dataset was used to run the ABGD algorithm online (http://wwwabi.snv.jussieu.fr/public/abgd/abgdweb.html), using the default priors (Pmin = 0.001, Pmax = 0.1, X = 1.5). These prior settings are optimized for use with the barcoding gene COI. Since the ABGD algorithm assumes the input sequences to belong to barcoding genes, only COI was used in this analysis.

4. Bayesian species delimitation

As opposed to the abovementioned methods for species discovery, BP&P (Yang & Rannala, 2010) requires an initial allocation of sequences to putative species and a species tree as input. The algorithm then compares different models of species delimitation, in which each model is generated by lumping sets of initially defined species together. The algorithm cannot split up putative species, and is therefore often termed species validation algorithm, as opposed to the other methods in this study, which can be seen as species discovery methods. Here, we use BP&P to estimate the posterior speciation probabilities for the 28 putative species derived from the GMYC analysis (for the COI region, Yule tree prior). This hypothesis concerning species boundaries contained most of the splits within morphotypes consistently recovered across methods. The split of *E. descendens* was not included in the BP&P analysis, since this would generate species consisting of a single specimen. A sequential approach, combining methods for species discovery (such as GMYC) and an algorithm for species validation (such as BP&P) has been advocated by several authors (e.g. Carstens et al., 2013; Barrett & Freudstein 2011; Setiadi et al., 2011). Apart from an initial species allocation and species tree, several other parameters influence the BP&P analysis: prior distributions for τ and θ, and utilization of algorithm 1 or 0. We followed Leaché et al. (2010), running the analysis with three different combinations of prior distributions for τ and θ. First, a prior combination assuming relatively small ancestral population sizes: $\theta \sim (2,2000)$ and shallow divergences among putative species: $\tau \sim (2,2000)$. Second, a prior combination assuming relatively large ancestral populations: $\theta \sim (1,10)$ and deep divergences: $\tau \sim (1,10)$. Finally, a conservative

combination of priors favoring the lumping of taxa, with large ancestral populations: θ ~ (1,10), and shallow divergence times: τ ~ (2,2000). Each of these prior combinations was run under both algorithm 0 and 1. The five molecular markers were analyzed as independently evolving loci. The species tree used in this analysis is the maximum clade credibility tree based on the *BEAST analysis of the 5 marker dataset

Results: Higher phylogenetic relationships

None of the inferred gene trees showed fully resolved evolutionary relationships within *Eristalinus* s.l.. Resolution was especially low for the nuclear genes (18S, 28S) and the mitochondrial 12S. Nevertheless, Bayesian inference recovered support for a limited number of clades coinciding with previously described morphospecies. Gene trees based on the remaining two mitochondrial regions (COI and *cyt* b) provided support for several of the relationships among putative species, but left the deeper nodes largely unresolved. Of the subgenera defined by Thompson (2003), only the subgenus *Eristalodes* is supported as monophyletic in the COI gene tree (PP: 0.98). One previous molecular study (Pérez-Bañón et al., 2003), recovered a sister relationship between a representative of *Eristalodes* (*E. taeniops*) and *E. megacepalus* (*Eristalinus* s.s.). These findings were further corroborated by the morphology of male genitalia, urging the authors to argue against the taxonomical importance of eye ornamentation in *Eristalinus* s.l.. Our findings, however, combined with the fact that sampling by Pérez-Bañón et al. (2003) only included a single taxon with fasciate eye ornamentation, suggest that *Eristalodes* as traditionally described (with fasciate eye ornamentation) represents a monophyletic, coherent evolutionary group. Nonetheless, recognition of this taxon at the genus level renders the genus *Eristalinus* paraphyletic. Resolution of this classification must however await further studies providing higher resolution for the deeper nodes in the *Eristalinus* s.l. evolutionary history. The current lack of this resolution could be due to bouts of rapid diversification during the early history of the group, as suggested in other groups (Psonis et al., 2017; Shaw, 2002), causing low phylogenetic signal for this part of the tree. With the current dataset however, this hypothesis remains untestable, requiring evidence from additional molecular markers. One potential method to probe into deeper evolutionary relationships within the genus is provided by the sequencing of mitogenomes. The utilization of parallel sequencing to obtain complete mitochondrial genomic sequences to solve deeper nodes in rapid radiations has been successfully tested in other groups (White et al., 2011; Botero-Castro et al., 2013; Wu et al., 2014; Li et al., 2016; Raupach et al., 2016).

Results: Species delimitation in *Eristalinus* s.l.:

The number of putative species recovered by the analyses performed in this study ranged from 19 (ABGD algorithm using the COI dataset) to 58 (GMYC algorithm, for the *cyt* b dataset, multiple thresholds and Yule tree) (Table 1; Figure 4). Only three of these clusters are shared among all analyses, and coincide with the nominal taxa *Eristalinus monozonus*, *E. gymnops* and one undescribed morphotype. This agreement among algorithms, supported by their uniform morphology, strongly supports the recognition of these clusters as species. The undescribed morphotype will be formally described elsewhere, including a full morphological diagnose.

method	GMYC	GMYC	GMYC	GMYC	GMYC	GMYC	GMYC	GMYC	PTP	PTP	ABGD	ABGD	ABGD	ABGD
marker	COI	COI	cytb	cytb	COI	COI	cytb	cytb	COI	cytb	COI	COI	COI	COI
parameters	single_C	single_Y	single_C	single_Y	multiple_C	multiple_Y	multiple_C	multiple_Y			base	4,64	2,78	7,74
E. descendens	S2	R	S2	S2	S2	S2	S2	S2	S2	S2	S2	S2	S2	S2
E. modestus	S2	S3	S2	S3	S2	R	S3	S5	S3	S2	R	S2	R	R
E. gymnops	R	R	R	R	R	R	R	R	R	R	R	R	R	R
Eristalinus sp. nov.1	R	R	R	R	R	R	S2	R	R	R	R	R	R	R
E. tabanoides	S2	S2	S3	S3	S2	S3	S5	S4	S2	S2	R	R	S2	R
E. monozonus	R	R	R	R	R	R	R	R	R	R	R	R	R	R
E. eclara	R	R	R	R	R	S2	S2	S2	R	R	R	R	R	R
E. cupreus	R	R	R	R	R	R	S2	R	R	R	R	R	R	R
E. taeniops	R	R	R	R	R	R	S2	R	R	S2	R	R	R	R
Eristalinus sp. nov.2	R	R	R	R	R	R	R	R	R	R	R	R	R	R
E. cressoni	S2	S2	R	R	S2	S2	R	R	S2	R	R	R	R	R
E. haplops	R	R	S7	S5	R	R	S6	S7	R	S7	R	R	R	R
E. smaragdinus	R	R	R	R	R	S2	R	S2	R	R	R	R	R	R
E. lineifacies	S2	S2	S3	S3	S4	S4	S6	S4	S2	S3	R	R	R	R
E. cacops	R	R	R	R	R	R	S2	R	R	R	R	R	R	R
E. vicarians	S4	S3	S3	S3	S2	S3	S4	S6	S8	S4	R	R	S4	R
E. virescens	R	R	R	R	R	S2	R	R	R	S2	R	R	R	R
E. madagascariensis	R	R	R	R	R	R	R	R	R	R	R	R	L	R
E. fuscicornis	R	R	S2	S2	R	S3	S3	S2	R	S2	L	L	L	
E. quinquelineatus	-	-	-	-	-	-	-	-	-	-	-	-		
E. barclayi	-	-	-	-	-	-	-	-	-	-	-	-	-	-
singleton1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
singleton2	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Clusters	25	26	26	27	25	30	30	31	23	25	19	21	25	20
Entities	31	29	40	37	35	42	57	57	35	42	23	24	32	24
Singletons	6	3	14	10	10	12	27	26	12	17	4	3	7	4

Table 1: Summary of the relationship between the nominal taxa (morphotypes) used in this study and the various species delimitation methods. The top three rows give the method, marker and parameters (Y: Yule tree prior, C: coalescent tree prior, single: single threshold, multiple: multiple threshold, base: baseline, numbers give prior maximal distances in ABGD analyses) used in each analysis. Morphotypes can either be recovered as a single cluster (R), split up into X smaller clusters (SX), or lumped together with specimens belonging to another morphotype (L). Owing to the higher complexity, associations for a subset of morphotypes contained in subgenus *Eristalodes* are depicted separately in Figure 4.

For the remaining morphospecies, clusters inferred by the molecular delimitation methods exhibit varying degrees of disagreement. With the exception of the subgenus *Eristalodes* (and *E. vicarians* in one analysis), these disagreements are nevertheless limited to differential splitting of morphospecies, rendering all resulting clusters morphologically homogenous. Since the majority of these clusters are not shared among delimitation algorithms, their biological significance remains unsupported. Splits coinciding with nominal taxa are however recovered in all analyses, providing support for their recognition as species.

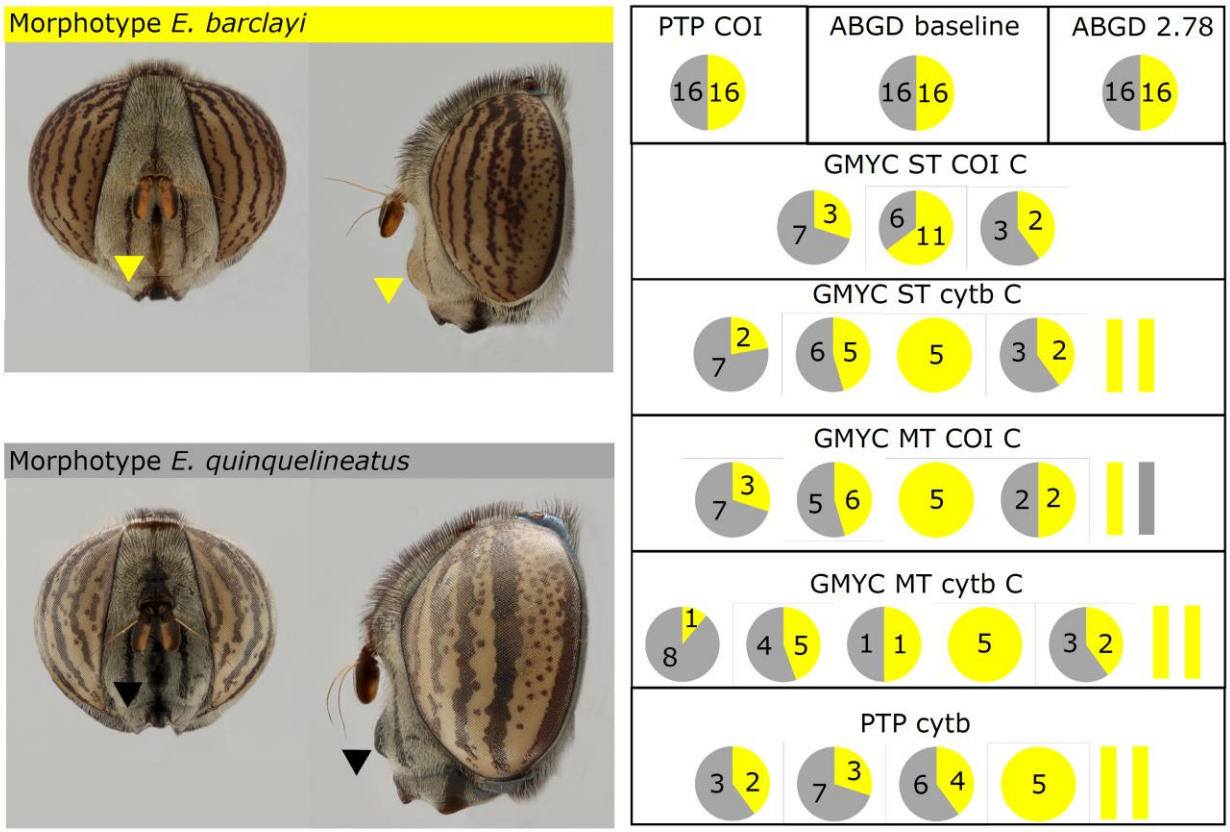

Figure 4: Relationship between morphological identification and cluster recovered in different molecular-based species delimitation algorithms in the subgenus *Eristalodes*. Clusters recovered in each species delimitation analysis are represented as circles, with the number of specimens identified as *Eristalinus quinquelineatus* indicated in gray while those corresponds to morphotype *E. barclayi* are indicated in yellow. Vertical lines represent singletons.

The multitude of smaller clusters found within morphospecies can in most cases be attributed to possible artefacts in the dataset and/or methods used. Firstly, the multi-threshold version of the GMYC analysis consistently recovers more clusters than the single-threshold counterpart. This observation is in line with those made by Fujisawa & Barraclough (2013), who found that the multi-threshold version of GMYC yields fairly similar results to the single-threshold version, but has a tendency to over-split. Secondly, the *cyt*b dataset recovers more clusters in both the GMYC and PTP analyses compared to the COI dataset. Higher variability of the former marker, as reflected in branch lengths and resolution of the gene tree can explain these differences. In this, the current dataset emphasizes the importance of sequencing multiple loci to account for discordant evolutionary histories and mutation rates across markers. However, despite these possible artefacts, several splits are consistently recovered across analyses, recovering genetic structure inside

morphologically homogenous groups: *E. vicarians*, *E. descendens*, *E. lineifacies*, *E. tabanoides* and *E. modestus*. Furthermore, the split occurring in the latter morphotype is supported by all coalescent-based species validation analyses (BP&P), providing strong indications that this morphologically homogenous groups could contain independent evolutionary lines, worthy of description at the species level.

Of the remaining morphospecies exhibiting consistent splits across methods, splits receive high support from two out of three prior combinations for *E. vicarians* and *E. tabanoides*, while the third prior combination is marginally unsupported (0.93 for *E. tabanoiodes*) or unsupported (0.88 for *E. vicarians*). We would therefore suggest that these morphospecies could potentially contain multiple independent evolutionary lines, and should be subjected to in-depth and population level scrutiny. The remaining splits on the other hand, do not receive support from BP&P suggesting these lineages to not represent independent evolutionary lineages. They could, however, represent incipient speciation, cryptic species (e.g. Hebert et al., 2004; Bickford et al., 2007; Ramaiya et al., 2010; Pramual & Kuvangkadilok, 2012; Kuchta et al., 2016), or geographic structuring of populations (e.g. Welton et al., 2013). The split observed in the morphologically homogenous group of specimens initially identified as E. lineifacies, can be explained by geographic origin of the samples. One of the clusters was collected in Ethiopia, whilst the second group of specimens occurs in South Africa. In order to confirm the presence of incipient speciation in this group, representatives from the intervening geographical areas should be sampled. The evidence presented here can best be interpreted as the presence of molecular divergence in a coherent species, caused by geographical isolation. For the other morphospecies exhibiting consistent splits across analyses, no clear correlation was found between these clusters and geographical or morphological characteristics. Nevertheless, these splits could indicate an ongoing diversification within these nominal taxa, which could be explored further by increasing the sampling to include their full geographic range.

Considering the abovementioned artefacts, and consistent splits among morphospecies, we suggest the recognition of all nominal taxa of *Eristalinus* s.s. (excluding *Eristalodes*; see Figure 4) present in this study on the species level. Furthermore, sufficient support is presented in this study to recognize the cluster labeled *Eristalinus* sp. nov. 1 as a new, morphologically divergent species.

Morphological characters used to distinguish nominal taxa throughout the genus *Eristalinus* (e.g. colour and shape of metafemur, abdominal banding pattern, ...) are conserved within subgenus *Eristalodes*. Therefore, described taxa are identified using subtle morphological characters, such as colour of basoflagellomeres, antennae and facial tubercle (Figure 4). Of the six morphospecies identified in our dataset, one showed a combination of characters not matching any presently described taxon (*Eristalinus* sp. nov. 2). Furthermore, this putative new species, as well as three described taxa (*E. cressoni*, *E. taeniops* and *E. fuscicornis*) are well-supported by the majority of the molecular sequence delimitation algorithms and the species validation algorithm BP&P. In contrast, the remaining two morphospecies (*E. quinquelineatus* and *E. barclayi*) are not consistently supported by these algorithms (Figure 4). Moreover, clusters combining specimens showing

morphological characters ascribable to each of these morphospecies were recovered, bringing into question the diagnostic value of the morphological characters used to identify them. The limited morphological divergence coupled with the widespread disagreement among sequence-based algorithms hampers the recognition of these three nominal taxa as species. We therefore suggest the recognition of these nominal taxa as a species complex, possibly illustrating a recent divergence, ongoing speciation, hybridization or introgression.

Conclusions:
The incongruences observed between the various species delimitation methods highlight the importance of comparing results across algorithms and parameter sets. All contemporary species delimitation algorithms attempt to simplify the complex process of speciation by making a set of assumptions (Carstens et al., 2013). These assumptions, (e.g. no gene flow after divergence, absence of hybrids,...) however, could be violated in any empirical study system, rendering the results of the respective methods untrustworthy. Since many species delimitation studies focus on taxonomically difficult or understudied systems, testing the validity of assumptions such as absence of gene flow, panmictic populations, etc. becomes problematic. It is therefore hard to judge which algorithms have their assumptions met and will be more appropriate for the study system at hand. However, when various methods are compared and produce congruent species hypotheses, these delimitations gain confidence, especially if the utilized methods present complementary assumptions (Ence & Carstens, 2011; Carstens et al., 2013). Conflicts between methods, on the other hand, can be indicative of assumptions not being met or since speciation is mostly a gradual process (with the exception of allopolyploid speciation, see e.g. Kim et al., 2008), differences in sensitivity to detecting genetic divergence. These considerations are in line with the General Lineage Concept of species, were some operational criteria differ in their ability to discern between lineages still in the process of differentiation.

Unfortunately, the TAS approach using the nuclear molecular markers proved unsuccessful for the genus *Eristalinus*. As such we were unable to resolve the phylogenetic relationships among the different subgenera of *Eristalinus* and among the different species of the genus. In order to do so in the near future, however, we explored two NGS-techniques to sequence entire mitochondrial genomes. A first (PE methodology) was explored to obtain high-quality full mtDNA genomes that could be annotated for all gene (gene regions) and that will serve as a backbone for a larger full mtDNA sequencing protocol (HTS methodology).

Whole-mitochondrial genome sequencing (WmtGS): PE methodology

Another NGS technique that was explored within the SYRPINTINE project was the sequencing of entire mitochondrial genomes, which nowadays is becoming cheaper, especially since a high amount of data is usually obtained. Whole mtDNA genomes show a high phylogenetic resolution and may thus prove valuable to unravel the phylogenetic and taxonomic relationships within the genus *Eristalinus*. However, the full mtDNA genome of *Eristalinus* has only been sequenced for one, non-Afrotropical species. We thus sequenced the entire mitochondrial genome (mtDNA) using

protocols outlined in Tang et al. (2014, 2015; references given below) of five *Eristalinus* species, viz. *E. aeneus*, *E. barclayi*, *E. fuscicornis*, *E. tabanoides*, and *E. vicarians*.

Given the good success of the whole mtDNA sequencing, we further explored the technique for Syrphidae systematics, phylogeny and taxonomy. This has resulted in a collaboration with Jeff Skevington (Canadian National Collections of Insects, Arachnids and Nematodes, Toronta, Canada), Ximo Mengual (Zoologisches Forschungsmuseum Alexander Koenig, Bonn, Germany) and Alfried Vogler (Imperial College London, UK) to sequence additional species for their full mitochondrial genome. The technique used here is even less expensive than that used by BGI. Moreover, the technique allows sequencing mtDNA genomes from pooled extracts, as long as the mtDNAs of the different taxa are sufficiently diverged (i.e. > 5%). This means that we now will not only sequence the mtDNA genomes of the Eristalinus species, but also that of two other genera which are very abundant in Southern Benin and Southern Togo, viz. *Syritta* and *Mesembrius*. We thus may be able not only to solve the phylogeny and taxonomy of *Eristalinus* but also that of two other, potentially important pollinator genera

Despite growing insight into shallow-level evolutionary relationships within *Eristalinus* (Pérez-Bañón et al., 2003; De Smet et al. in prep.), deeper relationships remain unresolved, casting uncertainty on the evolutionary validity and genealogy of the currently recognized subgenera. This is especially relevant for the evolutionary placement of subgenus *Eristalodes*, which has been suggested to render the subgenus *Eristalinus* paraphyletic (see above). As evidenced in previous studies, sequencing of the markers traditionally used in Dipteran phylogenetics (Pérez-Bañon et al., 2003; De Smet et al. in prep.) is unable to provide the resolution for these genus-level relationships.

Phylogenetic studies utilizing whole mitochondrial genomes have shown the potential of these data to tackle phylogenetic issues at varying levels of classification (e.g. Cameron et al., 2007, 2009; Nelson et al., 2012; Ma et al., 2012; Cameron, 2014; Yong et al., 2015). Resolving the abovementioned issues in Eristalinus using mitogenomics has not been previously attempted, and no mitochondrial genomes for the genus are currently available on Genbank. The current study therefore aims to assemble and annotate the mitochondrial genome for five *Eristalinus* species, from two subgenera, viz. subgenus *Eristalinus* s.s.: (*Eristalinus aeneus*, *Eristalinus vicarians*, *Eristalinus tabanoides*) and the subgenus *Eristalodes* (*Eristalinus barclayi*, *Eristalinus fuscicornis*).

Specimen and genomic DNA extraction:
Total genomic DNA was extracted from five specimens (Figure 5; Table 2) using the DNeasy® Blood &Tissue kit (Qiagen Inc., Hilden, Germany). To minimize contamination with bacterial DNA, the extraction was limited to three legs for each specimen. Quantification of DNA extracts was performed with Qubit 2.0 (Invitrogen, Life technologies).

Table 2 Specimens for which the entire mtDNA mitogenome was sequenced.

YDS2	*Eristalinus tabanoides*	male	Benin	Calavi	April 2014
YDS5	*Eristalinus vicarians*	female	Benin	Calavi	April 2014
YDS8	*Eristalinus aeneus*	female	Ethiopia	Holeta	October 2012
YDS3	*Eristalinus fuscicornis*	female	Benin	Cotonou	unknown

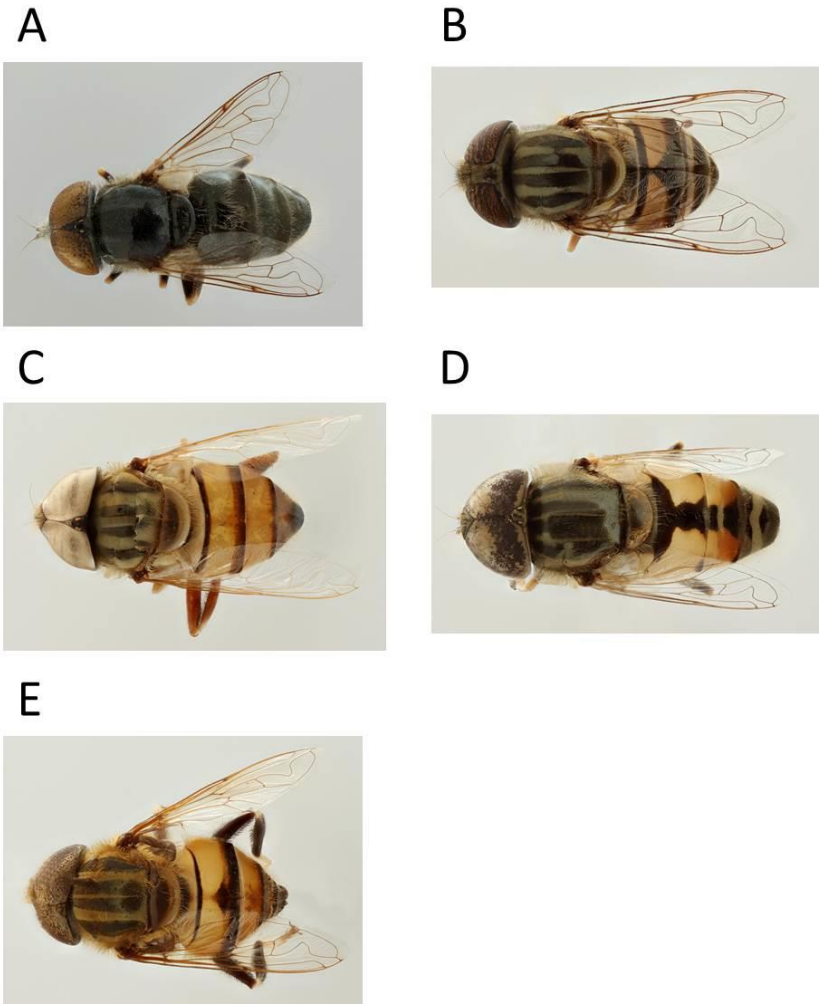

Figure 5: The five Eristalinus species of which the entire mtDNA mitogenome was sequenced: A. *Eristalinus (Eristalinus) aeneus*; B. *Eristalinus (Eristalodes) barclayi*; C. *Eristalinus (Eristalodes) fuscicornis*; D; *Eristalinus (Eristalinus) tabanoides*; E. *Eristalinus (Eristalinus) vicarians*.

Genome sequencing and analysis:

For each specimen, 100ng of DNA was pooled in a 250bp insert Hiseq DNA library, following manufacturer's instruction. The library was sequenced using the 150bp PE method on an Illumina Hiseq4000 platform at BGI-Shenzen, China. Using a custom Perl script (Zhou et al., 2013), reads containing adaptor contamination, > 10% low quality bases (Phre < 20) or > 5 unresolved bases (N's) were removed from the raw data. The remaining high quality reads were submitted to the mitogenome assembly pipeline described by Tang et al. (2014) for de novo assembly, using

SOAPdenovo-Trans (-K 71, -t 1) for assembly of reads, and *Simosyrphus grandicornis* (GenBank accession DQ866050.1) as a reference. Two parallel approaches were used in the annotation of the resulting mitogenomes, with the aim of comparing results. The first approach utilized the default settings of MITOS (Bernt et al., 2013) and the invertebrate genetic code for mitochondria. Secondly, the mitogenomes were annotated manually. For this all five mitogenomes were aligned against six dipteran mitogenomes obtained from GenBank (*Simosyrphus grandicornis* NC008754; *Nemopoda mamaevi* KM605250; *Zeugodacus caudatis* KT625491; *Zeugodacus caudatus* KT625492; *Culex tritaeniorhynchus* KT852976; *Culex pipiens* subsp. *pallens* KT851543), and ORF for the respective mitochondrial genes were inferred using the tRNA punctuation principle as proposed by Cameron (2014). Cloverleaf structure for the 22 inferred tRNAs were visualized in MITOS, circular genomes were reconstructed and visualized in Geneious 10.2.2 (Biomatters, New Zealand).

Genetic divergence:
Mitochondrial genomes were split into three separate datasets for analyzing genetic distance between sampled species. The first and second dataset contained only rRNA and protein coding sequences respectively, while the final dataset contained both rRNA and PCG. Alignments were generated using MUSCLE (Edgar, 2004), and uncorrected p-distances between *Eristalinus* species pairs were calculated in Geneious.

General features:
Complete mitochondrial genome sequences recovered in this study ranged from 15,757bp (*E. barclayi*) to 16,245bp (*E. aeneus*) (Figure 6, Table 3). This variation can be largely ascribed to varying success in generating reliable sequence data for the A + T-rich control area (Table 3). Gene order was similar across *Eristalinus* specimens and in line with previously reported dipteran mitochondrial genomes (e.g. Pu et al., 2017; Yong et al., 2015; Li et al., 2017). Every sampled individual contained the same 37 genes, of which 13 were protein-coding genes (PCG), 2 rRNA genes and 22 tRNA genes (Figure 6). The non-coding control region was highly variably across individuals, measuring 1263bp in *E. aeneus*, 735bp in *E. tabanoides*, 976bp in *E. vicarians*, 839bp in *E. barclayi* and 890bp in *E. fuscicornis*. In all specimens, this region was flanked by the trnS and trnI genes. Furthermore, the number of intergenic regions, and their cumulative length was highly variable across sampled species. Three instances of overlap were observed between adjacent genes on the same strand (only two for *E. aeneus* and *E. tabanoides*), three between adjacent genes on opposite strands for all species but *E. aeneus*. Base composition of *Eristalinus* mitochondrial genomes is heavily skewed towards A-T (% AT > % GC), with a GC content of 20% for most genomes recovered in this study (Table 3).

Table 3: Genome size (bp), base composition (%) and GC content (%) for the mitochondrial genomes of the five Syrphidae.

	Genome size	A	T	GC	GC content	
Eristalinus tabanoides	15.792	41.2	38.8	8.3	11.7	20.0
Eristalinus vicarians	15.966	41.1	38.9	8.2	11.8	20.0
Eristalinus aeneus	16.245	47.8	32.0	7.1	13.1	20.2
Eristalodes fuscicornis	15.815	41.0	38.9	8.4	11.7	20.1
Eristalodes barclayi	15.757	40.9	39.0	8.4	11.8	20.2

Protein-coding genes:

Of the 13 protein coding genes identified in *Eristalinus* mitochondrial genomes, nine were situated on the J-strand, while the remaining 4 were situated on the N-strand. Stop and start codons and the length of the coding regions are given in Table 3. The respective PCG's had identical start and stop codons across the sampled species, with the exception of nad6 in E. aeneus which had an ATT start codon instead of ATC and nad1, which had a TAA stop codon in *E. aeneus*, *E. barclayi* and *E. fuscicornis*, while having TAG in *E. tabanoides* and *E. vicarians*. The most common start codon was ATG with 6 occurrences, while ATC and ATT were the second most prevalent with 2 occurrences (however, in *E. aeneus* they occurred 1 and 3 times, respectively). The remaining start codons (TCG, TTG, GTG) each occurred only once. The most common stop codon was TAA, which was found in 7 genes, while the remaining PCG's had incomplete stop codons.

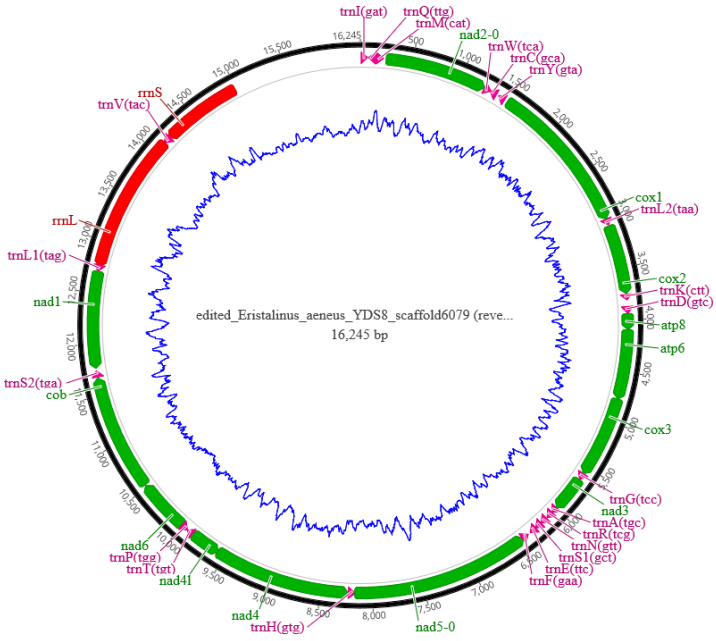

Figure 6: Complete mitogenome of the hover fly *Eristalinus barclayi* with BRIG visualization showing the protein coding genes, rRNAs, tRNAs and non-coding regions.

Transfer and ribosomal RNA genes:

Of the 22 tRNA genes identified in the mitochondrial genomes (Figures 6-7), 14 were situated on the J-strand, while the rest was situated on the N-strand. The cloverleaf structure for trnS1 lacked the D-loop in all species (Figure 6), while trnK and trnR lacked the TΨC-loop in *E. aeneus* and *E. tabanoides*, respectively. Both recovered rRNA genes were situated on the N-strand for all five *Eristalinus* species.

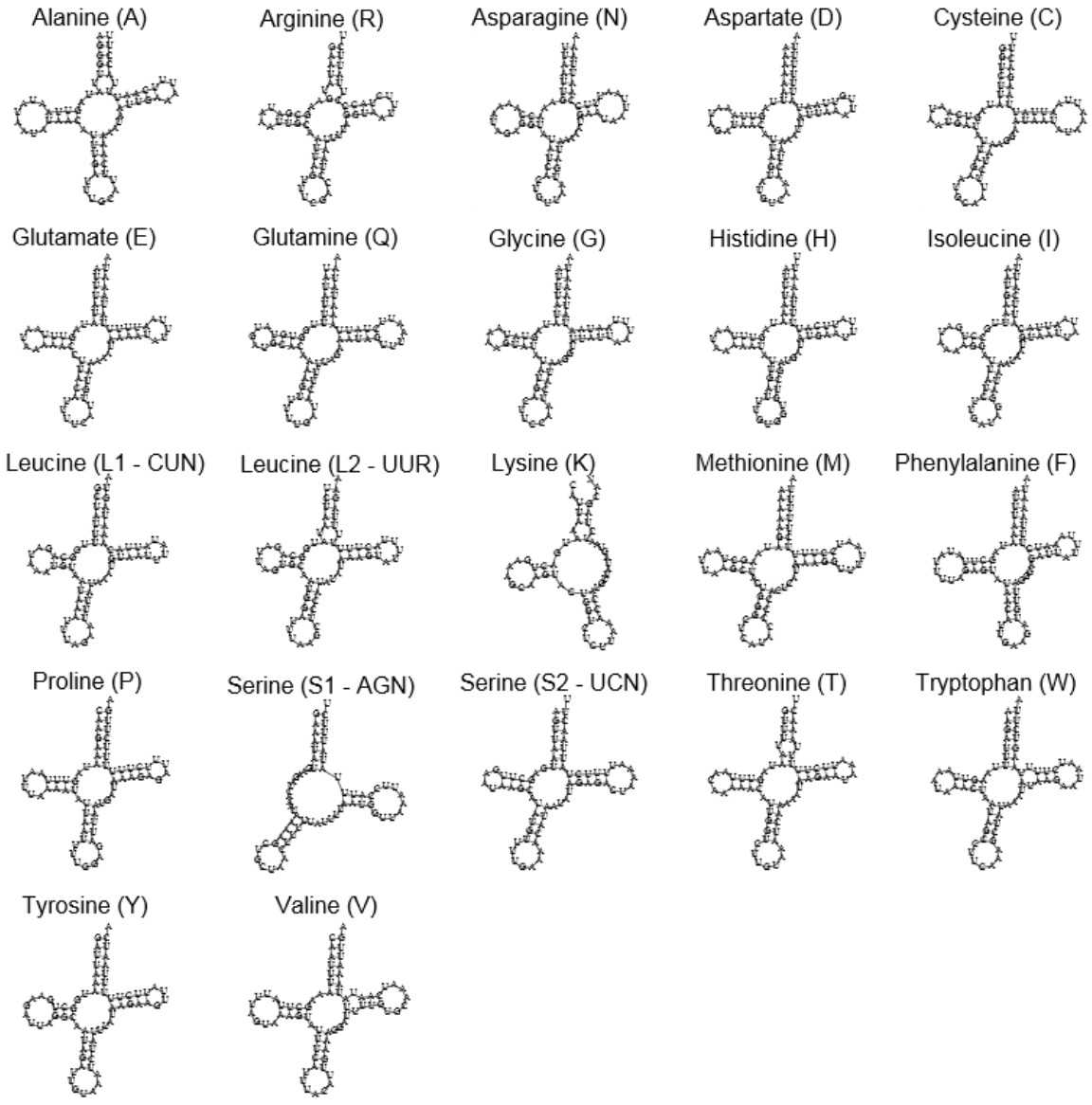

Figure 7: Cloverleaf structure of the 22 inferred tRNAs in the mitogenome of the hover fly *Eristalinus aeneus*.

This study presents the first annotated mitochondrial genomes for the hover fly genus *Eristalinus*. Within the family Syrphidae, the only other mitochondrial genomes available are: *Eristalis tenax* (16,091bp, Li et al., 2017) *Episyrphus balteatus* and *Eupeodes corollae* (16,175bp and 15326bp respectively, Pu et al., 2017), *Ocyptamus sativus* (partial genome) and *Simosyrphus grandicornis* (15,214bp and 16,141bp, Junqueira et al., 2016) and the partial genome for an unknown genus

"Syrphidae sp." (11,583bp, Tang et al., 2014). Genome sizes recovered in this study are comparable with those already published for the family (Table 2), with *E. aeneus* showing the largest published Syrphidae genome to date (16,245bp). Gene order is identical across the sampled *Eristalinus* species, representing two subgenera of the genus (*Eristalodes* and *Eristalinus*), and concurs with that found in other Syrphidae.

Comparing manual (using annotated Genbank sequences) to automated (MITOS, Bernt et al., 2013) annotation highlighted several conflicts owing to specific properties of mitochondrial genomes. Especially, the presence of overlap between adjacent open reading frames (Table 3) and truncated stop codons (completed by RNA polyadenylation) caused uncertainty regarding termination of genes. In some cases, two alternative stop codon positions for an open reading frame could be proposed, one theorizing an overlap between adjacent ORF's, while the alternative would involve an interrupted stop codon, missing the last A nucleotides of a TAA stop codon. In these cases, following the punctuation method proposed by Cameron (2014), preference was given to introducing a truncated stop codon.

The utility of mitochondrial genomics for the study of insect evolution and phylogeny has been amply demonstrated (Cameron, 2013). Mitochondrial genomes are available for all Insect orders, and have been shown to hold phylogenetic information for inference over extensive taxonomic scales (e.g. Cameron et al., 2007, 2009; Nelson et al., 2012; Zhao et al., 2013; Ma et al., 2012; Logue et al., 2013). The mitochondrial genomes obtained in this study show a promising array of diversity across the genus *Eristalinus*, with low pairwise genetic distance between the two representatives of the subgenus *Eristalodes*. The presence of mitochondrial genomic diversity at this level, and the availability of mitochondrial genomes of close (*E. tenax*) and more distantly related taxa suggest that mitochondrial phylogenetics could provide insight into the deeper evolutionary relationships within *Eristalinus*, which up until now have remained unresolved. Moreover, the 5 full mtDNA genomes of the genus *Eristalinus* will serve as a backbone mtDNA genome, not only to resolve the phylogenetic relationships within the genus *Eristalinus*, but also that of other genera of Syrphidae of which the various species are involved in plant pollination. This will allow us to better disentangle the diversity of plant pollinators involved in the plant-pollinator networks in Benin, and hopefully beyond. For this, we have started a high-throughput shotgun sequencing (HTS) approach to resolve the species boundaries and phylogenetic relationships of the hover fly genera *Eristalinus*, *Eristalis*, *Mesembrius*, *Phytomia*, *Chasmomma*, and *Syritta* of which the members are often found on flowers.

DAP Dual-

Whole-mitochondrial genome sequencing (WmtGS): HTS methodology

High-throughput shotgun sequencing (HTS) of mixtures of specimens and the bioinformatic reconstruction of mitochondrial genomes (mitochondrial metagenomics or mitometagenomics, MMG) has been successfully applied to biodiversity inventories (Tang et al. 2014; Crampton-Platt et al. 2015; Gómez-Rodríguez et al. 2015). The methodology allows the rapid identification of the biological entities present in a complex sample. This method is a 'genome skimming' (Straub et al. 2012) approach, taking advantage of the fact that the mitochondrial genome is present in numerous copies per nuclear genome and thus is amenable to de novo genomic assembly at fairly low sequencing depth (Zhou et al. 2013; Tang et al. 2014; Crampton-Platt et al. 2015).

In the present study, this method of sequencing multiple mitochondrial genomes from a "unknown" mixed sample of individuals has been adapted to obtain the mitochondrial genomes of a taxonomically known set of samples. This is realized through barcoding each individual prior to pooling and sequencing, using the Illumina TruSeq Nano library preparation kit, allowing the pooling of up to 96 individually barcoded individuals.

Current progress: Morphological identification have been performed and verified by Kurt Jordaens and Yannick De Smet. DNA was extracted and the DNA quality was checked for 136 syrphid vouchers (genera *Eristalinus*, *Eristalis*, *Syritta*, *Chasmomma*, *Mesembrius*, *Phytomia*). Genomic libraries were prepared for 65 individually indexed vouchers (Table 4) using Ilumina TruSeq DNA Nano library preparation kit according to manufacturer's instructions (testing Low and High Sample Input mode for different specimens). The Library prep included:
- DNA fragmentation to +/- 350bp on a Covaris platform (75 sec at default settings)
- Fragment blunt ending and size selection with sample purification beads
- A-tailing and dual index ligation

The completed libraries were checked for quality and concentration on Bioanalyzer and Qubit, and pooled into a single sample. The resulting pooled libraries were sent to Macrogen for WGS on an Illumina MiSeq 2 x 300bp platform (X lanes, Y Gb output, currently waiting for the raw reads). We expect to have the raw results by the end of March. Individual mtDNA genomes will then be constructed using the 5 full mtDNA genomes obtained with the PE methodology. Finally, full mtDNA genomes will then be used to reconstruct the phylogeny of the different genera. In addition, we will be able to develop PCR primers for short DNA fragments that can be used to identify the different species using a classical, low-cost DNA barcoding approach.

Table 4 (next page): The 65 individual of the hover fly genera *Eristalinus*, *Eristalis*, *Syritta*, *Mesembrius* and *Xylota* for which genomic libraries were prepared indicating the DAP Dual-indexed layout (indexes).

		DAP Dual-Indexed Layout	AB-number	genus	species	sample name
A1	ATTACTCG-TATAGCCT	A1	AB59841346	*Eristalinus*	*ampyx1*	E_ampAB59841346
A2	TCCGGAGA-TATAGCCT	A2	AB59850621	*Eristalinus*	*cfvicarians3*	E_cfvAB59850621
B2	TCCGGAGA-ATAGAGGC	B2	AB48964149	*Eristalinus*	*cressoni*	E_creAB48964149
A3	CGCTCATT-TATAGCCT	A3	AB59907219	*Eristalinus*	*haplops*	E_hapAB59907219
D3	CGCTCATT-GGCTCTGA	D3	AB59907215	*Eristalinus*	*mendax*	E_menAB59907215
F3	CGCTCATT-TAATCTTA	F3	AB59940827	*Eristalinus*	*myiatropinus*	E_myiAB59940827
G3	CGCTCATT-CAGGACGT	G3	AB59907226	*Eristalinus*	*quinquelineatus1*	E_quiAB59907226
H3	CGCTCATT-GTACTGAC	H3	AB59841881	*Eristalinus*	*quinquelineatus2*	E_quiAB59841881
A4	GAGATTCC-TATAGCCT	A4	AB33593910	*Eristalinus*	*quinquelineatus3*	E_quiAB33593910
F4	GAGATTCC-TAATCTTA	F4	AB59907202	*Eristalinus*	*surcoufi*	E_surAB59907202
G4	GAGATTCC-CAGGACGT	G4	AB59907189	*Eristalinus*	*tabanoides*	E_tabAB59907189
H4	GAGATTCC-GTACTGAC	H4	AB59907209	*Eristalinus*	*taeniops*	E_taeAB59907209
A5	ATTCAGAA-TATAGCCT	A5	AB59938986	*Eristalinus*	*trizonatus*	E_triAB59938986
H5	ATTCAGAA-GTACTGAC	H5	AB49482393	*Eristalis*	*tenax*	E_tenAB49482393
H6	GAATTCGT-GTACTGAC	H6	AB49104188	*Mesembrius*	*chapini1*	M_chaAB49104188
A7	CTGAAGCT-TATAGCCT	A7	AB48985703	*Mesembrius*	*chapini2*	M_chaAB48985703
B7	CTGAAGCT-ATAGAGGC	B7	AB49134412	*Mesembrius*	*cyanipennis*	M_cyaAB49134412
C7	CTGAAGCT-CCTATCCT	C7	AB59870802	*Mesembrius*	*ingratus*	M_ingAB59870802
E7	CTGAAGCT-AGGCGAAG	E7	AB49104186	*Mesembrius*	*minor1*	M_minAB49104186
H7	CTGAAGCT-GTACTGAC	H7	AB59940799	*Mesembrius*	*sp2*	M_sp2AB59940799
A8	TAATGCGC-TATAGCCT	A8	AB59841868	*Mesembrius*	*sp3*	M_sp3AB59841868
B8	TAATGCGC-ATAGAGGC	B8	AB59870826	*Mesembrius*	*strigilatus*	M_strAB59870826
C8	TAATGCGC-CCTATCCT	C8	AB59313820	*Mesembrius*	*tarsatus*	M_tarAB59313820
E8	TAATGCGC-AGGCGAAG	E8	AB59841377	*Xylota*	*hancocki2*	X_hanAB59841377
H8	TAATGCGC-GTACTGAC	H8	AB49135813			_AB49135813
A9	CGGCTATG-TATAGCCT	A9	AB59841940	*Syritta*	*albopilosa*	S_albAB59841940
C9	CGGCTATG-CCTATCCT	C9	AB49104654	*Syritta*	*bulbus1*	S_bulAB49104654
D9	CGGCTATG-GGCTCTGA	D9	AB49104642	*Syritta*	*bulbus2*	S_bulAB49104642
E9	CGGCTATG-AGGCGAAG	E9	AB48985687	*Syritta*	*bulbus3*	S_bulAB48985687
F9	CGGCTATG-TAATCTTA	F9	AB49134401	*Syritta*	*bulbus4*	S_bulAB49134401
H9	CGGCTATG-GTACTGAC	H9	AB49478729	*Syritta*	*bulbus6*	S_bulAB49478729
A10	TCCGCGAA-TATAGCCT	A10	AB59841192	*Syritta*	*bulbus7*	S_bulAB59841192
B10	TCCGCGAA-ATAGAGGC	B10	AB59870807	*Syritta*	*bulbus8*	S_bulAB59870807
C10	TCCGCGAA-CCTATCCT	C10	AB59850595	*Syritta*	*bulbus9*	S_bulAB59850595

E10	TCCGCGAA-AGGCGAAG	E10	AB48933026	*Syritta*	*decora*	S_decAB48933026
F10	TCCGCGAA-TAATCTTA	F10	AB48985715	*Syritta*	*eminea*	S_emiAB48985715
H10	TCCGCGAA-GTACTGAC	H10	AB59841127	*Syritta*	*flaviventris2*	S_flaAB59841127
A11	TCTCGCGC-TATAGCCT	A11	AB59841221	*Syritta*	*flaviventris3*	S_flaAB59841221
B11	TCTCGCGC-ATAGAGGC	B11	AB48985707	*Syritta*	*hirta*	S_hirAB48985707
C11	TCTCGCGC-CCTATCCT	C11	AB59870825	*Syritta*	*hova*	S_hovAB59870825
D11	TCTCGCGC-GGCTCTGA	D11	AB48985731	*Syritta*	*leona*	S_leoAB48985731
E11	TCTCGCGC-AGGCGAAG	E11	AB59841097	*Syritta*	*leucopleura1*	S_leuAB59841097
F11	TCTCGCGC-TAATCTTA	F11	AB59940734	*Syritta*	*longiseta*	S_lonAB59940734
G11	TCTCGCGC-CAGGACGT	G11	AB48985712	*Syritta*	*similis*	S_simAB48985712
H11	TCTCGCGC-GTACTGAC	H11	AB49104647	*Syritta*	*sp1*	S_sp1AB49104647
A12	AGCGATAG-TATAGCCT	A12	AB49104641	*Syritta*	*sp2*	S_sp2AB49104641
B12	AGCGATAG-ATAGAGGC	B12	AB49104644	*Syritta*	*sp3*	S_sp3AB49104644
C12	AGCGATAG-CCTATCCT	C12	AB49135421	*Syritta*	*sp4*	S_sp4AB49135421
D12	AGCGATAG-GGCTCTGA	D12	AB49135433	*Syritta*	*sp5*	S_sp5AB49135433
C1	ATTACTCG-CCTATCCT	C1	AB59841208	*Eristalinus*	*ampyx2*	E_ampAB59841208
D1	ATTACTCG-GGCTCTGA	D1	AB49481107	*Eristalinus*	*andersoni*	E_andAB49481107
F1	ATTACTCG-TAATCTTA	F1	AB49481108	*Eristalinus*	*cf.longicornis*	E_cf.AB49481108
H1	ATTACTCG-GTACTGAC	H1	AB59850631	*Eristalinus*	*cf.vicarians2*	E_cf.AB59850631
E2	TCCGGAGA-AGGCGAAG	E2	AB48964149	*Eristalinus*	*cressoni*	E_creAB48964149
B5	ATTCAGAA-ATAGAGGC	B5	AB59881977	*Eristalinus*	*cupreus*	E_cupAB59881977
G5	ATTCAGAA-CAGGACGT	G5	AB49481114	*Eristalinus*	*euzonus*	E_euzAB49481114
C6	GAATTCGT-CCTATCCT	C6	AB49134428	*Eristalinus*	*flaveolus*	E_flaAB49134428
D6	GAATTCGT-GGCTCTGA	D6	AB49125024	*Eristalinus*	*fuscicornis1*	E_fusAB49125024
E6	GAATTCGT-AGGCGAAG	E6	AB49104160	*Eristalinus*	*fuscicornis2*	E_fusAB49104160
F6	GAATTCGT-TAATCTTA	F6	AB33597281	*Eristalinus*	*seychellarum*	E_seyAB33597281
G6	GAATTCGT-CAGGACGT	G6	AB59940532	*Eristalinus*	*smaragdinus2*	E_smaAB59940532
D7	CTGAAGCT-GGCTCTGA	D7	AB49134130	*Eristalinus*	*sp1*	E_sp1AB49134130
F7	CTGAAGCT-TAATCTTA	F7	AB59907209	*Eristalinus*	*taeniops*	E_taeAB59907209
D8	TAATGCGC-GGCTCTGA	D8	AB33597312	*Eristalinus*	*vicarians*	E_vicAB33597312
F8	TAATGCGC-TAATCTTA	F8	AB59870822	*Eristalis*	*apis*	E_apiAB59870822

Table 4: The 65 individual of the hover fly genera *Eristalinus*, *Eristalis*, *Syritta*, *Mesembrius* and *Xylota* for which genomic libraries were prepared indicating the DAP Dual-indexed layout (indexes).

Molecular analyses of plant material

In order to infer the pollination capacity of the *Eristalinus* species in Africa, pollen was examined using Scanning Electron Microscopy (SEM) images (Figure 8). Using SEM pictures a more detailed view of pollen ornamentation and ultrastructure was obtained than when solely using light microscopic imaging. Molecular methods were used to assess the correct identification of pollen that were linked with the local flora (Table 5) which was collected on site together with the *Eristalinus* hoverfly species. Sampling of flowering plant species was carried out simultaneously with that of the *Eristalinus* hoverflies to find out what the impact is of invasive or cultivated (non native) species on the pollination of the local flora.

Table 5: Morphological and TAS-based molecular identification of 28 flowering plant samples on which *Eristalinus* spp. hover flies were collected.

Plant ID n°	Morphological ID	Molecular ID (ITS)	Molecular (*matK*)	Molecular ID (*rbcL*)	Molecular (*trnH-psbA*)
1	***Palisota hirsuta***	*Palisota sp.*	*Palisota bracteosa*	*Palisota ambigua*	*Palisota sp.*
2	***Clerodendron inerme***	*Clerodendron sp.*	*Clerodendron splendens*	*Clerodendrum/Tetraclea*	*Clerodendron sp.*
3	***Clerodendron inerme***	*Clerodendron sp.*	*Clerodendron splendens*	*Clerodendrum/Tetraclea*	*Clerodendron sp.*
4	***Piper arboreum***	*Piper swartzianum*	*Piper sp.*	*Piper sp.*	*Piper arboreum*
5	***Chromolaena odorata***	*Chromolaena sp.*	*Chromolaena/Praxelis*	*Chromolaena/Praxelis*	*Chromolaena sp.*
6	***Sida rhombifolia***	*Sida sp.*	*Sida rhombifolia*	*Sida sp.*	*Sida rhombifolia*
7	**Combretaceae**	*/*	*/*	*/*	*/*
8	***Solanum torvum***	*Solanum sp.*	*Solanum torvum*	*Solanum chrysotrichum*	*Solanum torvum*
9	***Pluchea sp.***	*Pluchea sp.*	*Deinbollia/Pluchea*	*Deinbollia/Pluchea*	*Pluchea lanceolata*
10	***Ixora coccinea***	*Ixora guineensis*	*Ixora undulata*	*Ixora littoralis/guineensis*	*Ixora coccinea*
11	***Combretum molle***	*Combretum sp.*	*Combretum sp.*	*Combretum sp.*	*Combretum sp.*
12	***Echinacanthus attenuatus***	*Echinacanthus sp.*	*Echinacanthus sp.*	*Echinacanthus attenuatus*	*Echinacanthus sp.*
13	**Myrtaceae**	*/*	*/*	*/*	*/*
14	***Cyperus sp.***	*Cyperus sp.*	*Cyperus sp.*	*Cyperus sp.*	*Cyperus sp.*
15	***Alternanthera brasiliensis***	*Alternanthera brasiliensis*	*Alternanthera brasiliensis*	*Alternanthera sessilis*	*Alternanthera brasiliensis*
16	***Palisota hirsuta***	*Palisota sp.*	*Palisota bracteosa*	*Palisota ambigua*	*Palisota sp.*
17	***Impatiens irvingii***	*Impatiens pseudoviola*	*Impatiens balsamina*	*Impatiens walleriana*	*Impatiens walleriana*
18	***Coffea canephora***	*Coffea sp.*	*Coffea/Psilanthus sp.*	*Coffea/Psilanthus sp.*	*Coffea sp.*
19	***Alternanthera sessilis***	*Alternanthera halimifolia*	*Alternanthera halimifolia*	*Alteranthera ficoidea*	*Alternanthera halimifolia*
20	***Sida rhombifolia***	*Sida rhombifolia*	*Sida rhombifolia*	*Sida sp.*	*Sida rhombifolia*
21	***Ludwigia sp.***	*Ludwigia decurrens*	*Ludwigia sp.*	*Ludwigia octovalvis*	*Ludwigia sp.*
22	***Aneilema beniniense***	*Aneilema sp.*	*Aneilema sp.*	*Aneilema clarckei*	*Aneilema sp.*
23	***Gliricidia sepium***	*Gliricidia sepium*	*Gliricidia sepium*	*Robinia pseudoacacia*	*Gliricidia sepium*
24	***Mangifera indica***	*Mangifera indica*	*Mangifera indica*	*Mangifera indica*	*Mangifera indica*
25	***Terminalia mantaly***	*Terminalia sambesiaca*	*Terminalia complanata*	*Terminalia phellocarpa*	*Terminalia phanerophlebia*
26	***Rauvolfia vomitoria***	*Rauvolfia vomitoria*	*Rauvolfia sellowii*	*Rauvolfia vomitoria*	*Rauvolfia vomitoria*
27	***Syzygium guineense***	*Syzygium sp.*	*Syzygium sp.*	Myrtaceae	*Syzygium guineense*
28	***Caesalpina pulcherrima***	*Caesalpina pulcherrima*	*Caesalpina pulcherrima*	*Caesalpina pulcherrima*	*Caesalpina pulcherrima*

Leaf tissue sequencing:

In order to extract DNA from the leaf tissue, we improved our CTAB protocol so that it can also be used in the isolation of "ancient DNA". Prior to CTAB lysis, mucilaginous polysaccharides were removed using a washing buffer that contains 0.03 M N-laurylsarkosine and 0.13 M sorbitol. CTAB-

buffer contains 10 mM Tris, 20 mM EDTA, 0.02 M CTAB, 0.8 M NaCl, 3% PVP40. Every sample was kept for 60 minutes at 65°C in the CTAB lysis buffer. Chloroform extraction was carried out using a chloroform-isoamyl alcohol (24:1) mixture. Chloroform extraction is repeated three times. During the last step of the chloroform extraction, 2/3rd of the upper phase was pipetted to a different 1.5mL tube containing 99% ethanol and 3M Na-acetate. After centrifuging for 15 min at 13G at 6°C temperature, the pellet was washed with 70% ethanol, dried and dissolved in 100µL TE-buffer. In order to effectively remove additional proteins, polysaccharides, oligonucleotides and other contaminants for the isolated DNA, magnetic beads are used for DNA clean up.

DNA quantity of the isolated tissue samples is checked using Nanodrop and Qbit, whereas the DNA quality (length of fragments) is assessed using Fragment Analyzer. DNA barcodes were obtained following an approach of "target amplicon sequencing" (TAS). TAS is a rapid and cost effective approach that can be used to assess the genetic composition of an environmental sample but as well to quickly barcode a set of samples that need to be identified using barcoding markers. TAS utilizes a two-step PCR method that makes use of classic PCR followed by an amplification reaction that connects a labeled 10 bp tag to the amplified gene marker. This way, different samples can be distinguished from each other in the Illumina High Throughput Sequencing run. For our analyses on the identification of leaf tissue material, following gene makers (amplicons) were targeted: *matK*, *rbcL*, *trnH-psbA* and ITS (Table 5). Tagged amplicons were then directly sequenced on the Illumina platform of the Genomics Core (UZ-KULeuven). As such, this multiplex approach helped us in sequencing all necessary DNA barcoding fragments in one single run.

Using BarcodeCruncher, the large volume of data generated was handled. In addition, a bioinformatics pipeline was set up in the Botanic Garden to clean, trim, group, analyse, and simplify the obtained sequence data. A Blast search (as implemented in Geneious 10) was used to sort each read into groups according to species and barcode-gene.

Pollen analyses:
Complementary to the morphological analyses of the pollen on the *Eristalinus* hoverflies, the pollen of a small subset of the hoverflies was also investigated using molecular techniques. Pollen was removed from the *Eristalinus* hoverflies using a sterilized cotton swab and subsequently broken using a sonicator in order to remove the DNA from the pollen grains. DNA isolation of pollen samples was carried out using a slightly adapted protocol to that of the leaf tissue material. Compared to the leaf tissue DNA isolation protocol, the lysis buffer also contains proteinase K (0.2 mg, from a stock solution 20 mg/mL). DNA quantification and quality assessment is carried out in a similar way than with the leaf tissue material. HTS analyses are also carried out similarly to the protocol outlined above.
Pollen visualization – SEM (Figure 8):
Visualisation of pollen from flowers using Scanning Electron Microscopy was carried out as follows. Dried flowers or mature buds were rehydrated for 48h in Agepon® wetting agent (1:200, Agfa Gevaert, Leverkusen, Germany). Prior to critical point drying (CPD 030, Balzers), pollen material was dehydrated in a graded acetone series (30% - 50% - 70% - 100%). The critical point dried

anthers were mounted on stubs with double-sided adhesive tape and pollen grains were removed from the locules with a fine cactus needle to facilitate observation of the inner locule wall. Removed pollen grains were accumulated on the same stub for further observations. The stubs were sputter coated with gold (SPI-MODULETM Sputter Coater, SPI Supplies, West Chester, PA, USA). We used a Jeol JSM-6400 microscope, at 25 kV, for morphological observations.

In order to visualize the pollen that was present on the *Eristalinus* hoverflies we used a strategy that was based on the different pollination strategies of the plants that were collected on both compounds. Pollen presentation mechanisms differ between plants. Whereas some plants deposit their pollen on top of the head or thorax of the visiting insect (e.g. *Impatiens irvingii*), other plants deposit their pollen between the legs of the insects either where they are attached to the thorax or on the tarsi (e.g. *Terminalia mantaly*, *Mangifera indica*). In addition, pollen of some plants is eaten by insects to foresee in their protein intake and is therefore found on the different mouthparts (e.g. *Mangifera indica)*. *Eristalinus* hoverflies were critically point dried along with the pollen they carried (similar to the method used to analyse the pollen of the flowers collected). Rehydration treatment prior to critical point drying also follows the floral pollen rehydration strategy. Subsequent to critical point drying the *Eristalinus* hoverflies where dorsally attached to the stub so that the pollen on top of the insect could be visualized. Next the insects were detached from the stub and put on a new stub, this time ventrally so that the mouthparts and the legs can be visualized. Each time a new stub is used, the insect is sputter-coated with a thin layer of gold of only a few atoms thick (see above).

A Jeol JSM-6400 microscope was used for morphological observations at 25 kV. Observations took place at different magnifications in order to measure the overall size as well as to detect the different micromorphological characteristics that are specific for pollen of different species (e.g. aperture size, margo ornamentation, sexine sculpturing, presence/absence of Ubish bodies). Comparative size measurements of pollen were ascertained from digital SEM-micrographs using Carnoy 2.0 (Schols et al., 2002). Pollen terminology follows the Glossary of Pollen and Spore Terminology (Punt et al. 2007; http://www.bio.uu.nl/~palaeo/glossary/glos-int.htm). The terminology of pollen shape in polar view follows Reitsma (1970) while terms for shape classes in equatorial view were adopted from Erdtman (1971).

Results & Discussion:

Sequencing of the different amplicons for the collected plant species at IITA in Cotonou resulted in barcode sequences for each of the species (Table 5). Although barcoding is a useful tool to assign unidentified taxa to a genus, but identification to a species is less reliable. This is especially the case when dealing with species-rich clades. In addition, it is clear that a sufficient reference database is still lacking for the barcoding of African taxa in general. So despite new techniques evolving at an incredible pace, if no adequate database will be provided, genetic barcoding will never get at its full potential. Although the combination of four markers clearly improved the accuracy in species identification, there is still a significant percentage that could not be correctly identified using genetic barcodes. At genus level, a 100% match could be retrieved using *matK*, *trnH-psbA*, *rbcL* and ITS. When using for example only *rbcL* to delineate at genus level, only 80%

of the taxa could be correctly assigned to a certain genus. At species level, only 35% of the taxa could be correctly identified at species level using all four barcode markers. (Table 5). In case only *rbcL* was used to identify species, this proportion dropped to 12%.

Molecular analyses of pollen collected on *Eristalinus* hoverflies appeared to be less straightforward than with the leaf tissue samples. As a result, to date we have not been able to retrieve sequences from the pollen samples. DNA quantification analyses and quality assessment on the DNA retrieved from the pollen samples, demonstrated that the DNA contained only little contaminants. However, the amount of DNA was extremely low and appeared to be highly fragmented. The former is not really illogical when bearing mind that a pollen grains are single cells in most flowering plants containing a haploid gamete and that the amount of pollen found on the body of the insects was often rather low (see SEM palynology results). The latter issue is less obvious to explain however. In order to improve the sequencing of pollen grains, the current HTS protocol will be further improved so that also DNA of low quality and quantity can be sequenced.

SEM images of the pollen retrieved on different body parts of the *Eristalinus* hoverflies resulted in the detection of only 12 different pollen types (Figure 8). Based on morphological characters, 7 could be identified at species level, one at genus level and two at family level and one at order level. One pollen grain could not be identified due to the presence of only universal characters. Of the 11 different pollen grains that were observed, three occurred in more than 70 percent of the observations (*Sida rhombifolia*, *Mangifera indica* and *Terminalia mantaly*). Pollen of all three species were found on the legs and the thorax as well as on the different mouthparts demonstrating that the pollen of these three species is readily foraged on by the *Eristalinus* hoverflies. No pollen was found on the ventral side of the head or the thorax, meaning that local flora such as *Impatiens irvingii* was not visited by the hoverflies. Of the three species that were mainly visited, only one species is native to Benin (*Sida rhombifolia*), the other two are cultivated or planted as ornamental, showing that invasive plants can have a large impact on the fitness of the local flora as pollinator visits are seemingly drastically reduced. Other pollen that was found on the hoverflies mainly belonged to *Ixora coccinea* (Asian ornamental), *Piper arborescens* (Asian origin) and *Coffea canephora* (cultivated, origin tropical Africa), whereas only once we found a pollen grain that belonged to *Combretum molle* (local flora), *Aneilema beniniense* (local flora), an *Acacia* species and an unknown Asteraceae species (Figure 8). No other pollen grains were found on the hoverflies that matched with the list of species that were in flower when the *Eristalinus* hoverflies were collected (see above).

It is clear that non-native plants have a large impact on the visiting frequency of certain pollinators. When non-native flora causes pollinators to visit the local flora less abundantly than before their occurrence, this means that the fitness of the local flora is greatly impacted by the presence of new plant accessions, either planted as cultivar or ornamental, or as invasive weeds.

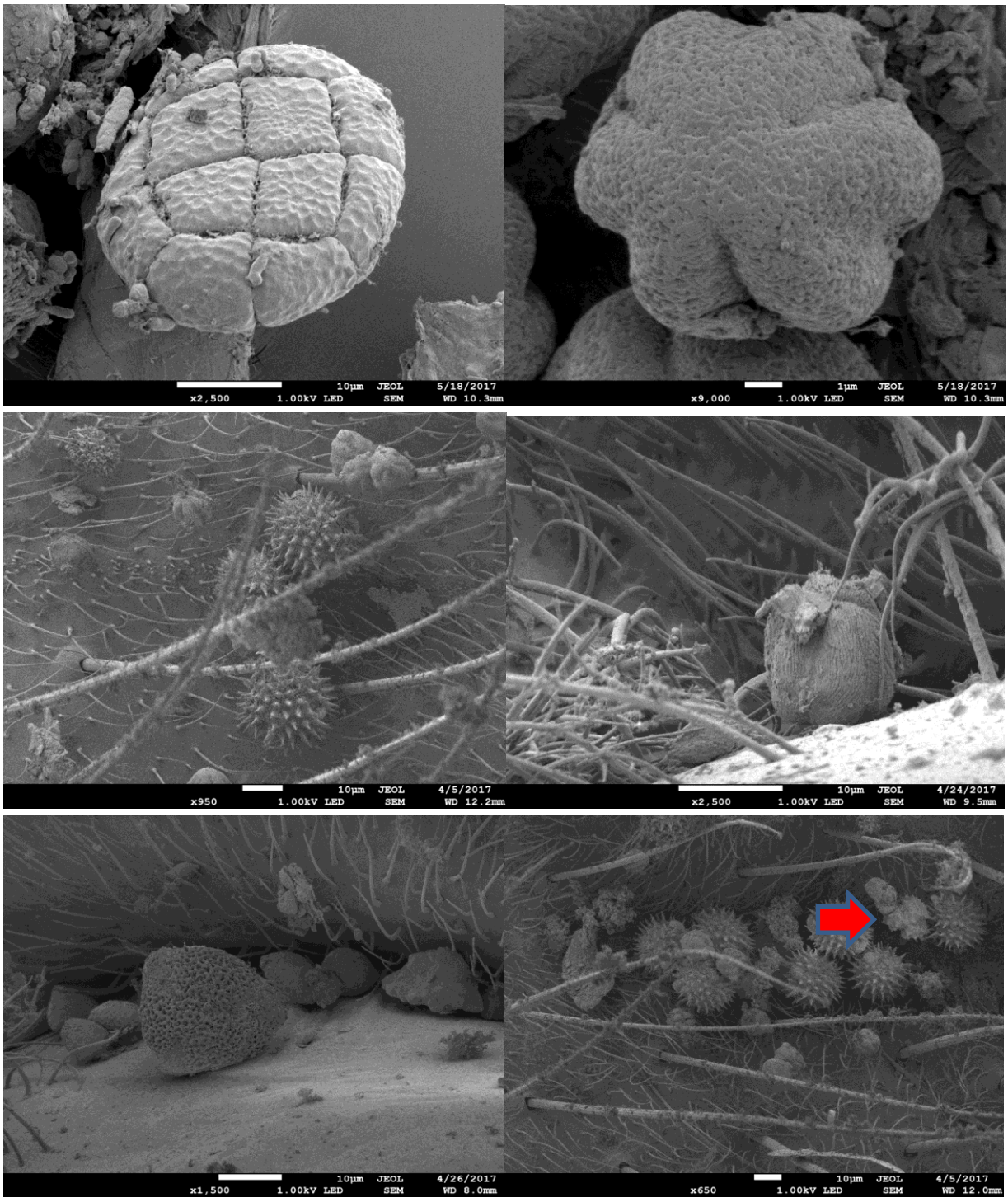

Figure 8: Pollen SEM images of pollen collected from *Eristalinus* sp. Top left: *Acacia* sp.; top right: *Terminalia mantely*; middle left: *Sida rhombifolia*; middle right: *Combretum molle*; down left: *Mangifera indica*; down right: Asteraceae sp. (red arrow).

3. DISSEMINATION AND VALORISATION

European Conference of Tropical Ecology (GTÖ) (Brussels, Belgium, 6-10th Feb. 2017)

Janssens S, Vanden Abeele S, Dauby G, Hawthorne W, Marschall C, Maurin O, Rejou-Mechain M, van der Bank M, Baya F, Beina D, Hardy O. 2017. How old is the African Flora? New insights in the evolution of African biomes based on large-scale dating and diversification analyses. European Conference of Tropical Ecology (GTÖ), Brussels, Belgium, 6-10 February 2017.

9th International Symposium on Syrphidae (Curitiba, Brazil, 28th Aug.-1st Sept. 2017).

Jordaens, K., Burt T. O., De Meyer, M., De Smet, Y., Goergen, G., Kelso, S., Mengual, X., Morales, M. N., Moran, K., Radenković, S., Reemer, M., Skevington J. H., Smit, J., Ssymank, A., Ståhls, G., van Steenis, J., Veličković, N., Vujić, A., Whittington, A. & A. D., Young, 2017. DNA barcoding of Afrotropical hoverflies (Diptera, Syrphidae): state-of-the-art, challenges and future perspectives. Book of abstracts of the 9th International Syrphidae Symposium, Curitiba, Brazil, 28th Aug.-1st Sept. 2017, p34-35.

Jordaens, K., Copeland, R., Goergen, G., Hamer, M., Janssens, S., Kirk-Spriggs, A. H., Midgley, J., Muller, B. & L., Njoroge, 2017. The Pollinator Information Network for Two-Winged Insects (PINDIP). Book of abstracts of the 9th International Syrphidae Symposium, Curitiba, Brazil, 28th Aug.-1st Sept. 2017, p36.

Jordaens, K., Goergen, G., Ssymank, A. & M., De Meyer, 2017. A preliminary checklist of the Syrphidae (Diptera) of Ghana, Togo, Benin and Nigeria. Book of abstracts of the 9th International Syrphidae Symposium, Curitiba, Brazil, 28th Aug.-1st Sept. 2017, p36.

Jordaens, K., Van Steenis, J., Goergen, G. & M., De Meyer, 2017. An improved taxonomy for the Afrotropical hover fly genus Syritta (Diptera, Syrphidae). Book of abstracts of the 9th International Syrphidae Symposium, Curitiba, Brazil, 28th Aug.-1st Sept. 2017, p37.

Jordaens, K., Goergen, G., Kelso, S., Skevington, J. H. & M., De Meyer, 2017. The Afrotropical hover fly genus Mesembrius (Diptera, Syrphidae) as an example of how DNA barcoding can improve the identification of sexually dimorphic insects. Book of abstracts of the 9th International Syrphidae Symposium, Curitiba, Brazil, 28th Aug.-1st Sept. 2017, p39.

7th International Barcode of Life Conference (Krugerpark, South Africa, 20-24 November 2017)

Jordaens, K., Goergen, G., Backeljau, T., De Meyer, M. & Meganck, K., 2017. The use of DNA barcoding to improve the taxonomy of Afrotropical hoverflies (Diptera: Syrphidae). Book of Abstracts, 7th International Barcode of Life Conference, Nov. 20-24, Kruger National Park, RSA.

Hardy, O., Dauby, G., Couvreur, T.L.P., Droissart, V., Rejou-Mechain, M., Janssens, S., Hawthorne, W., Marshall, C., Maurin, O., van der Bank, M., Thomas, D., Kenfack, D., Chuyong, G., Beina, D., Baya, F., 2017. The phylogenetic structure of plant assemblages in tropical Africa: from local

community to biogeographical scales. Book of Abstracts, 7th International Barcode of Life Conference, Nov. 20-24, Kruger National Park, RSA.

The SYRPINTINE project has now mostly focused on the molecular taxonomy of a selected number of genera of Afrotropical genera. These results have, and will further, shed light on the taxonomy and systematics of this important group of pollinators. Currently, we valorize the molecular results with a morphological study of the selected Syrphidae. At this moment, we are working on a first **identification key** for the Afrotropical *Syrphidae* of the genus Eristalinus and a revision of the identification key of the genus *Syritta*. To this end, Kurt Jordaens has received a RMCA Knowledge Management grant of eight months (April – December 2018) to hire a MSc who will assist in the production of online, freely-available identification keys for both genera. The person involved will produce series of stacking photography pictures for each species of both genera and develop the identification keys together with Yannick De Smet (*Eristalinus*) and Kurt Jordaens (*Syritta*). The distribution of identification keys will stimulate researchers in the Aftrotropical Region to include Syrphidae in their studies on plant-pollinator networks.

In light of this, Kurt Jordaens and Steven Janssens have received funding from the JRS Biodiversity Foundation (http://jrsbiodiversity.org/) to establish a **network on researchers and institutes involved in pollination biology in the Afrotropical Region**. This 3-year project (PINDIP: The Pollinator Information Network for two-winged Insects (Ditpera); see: http://jrsbiodiversity.org/grants/royal-museum-central-africa/) focusses on the dissemination of Diptera collection data and the relationship among Diptera pollinators and plants. The PINDIP project has its own website (https://www.pindip.org/) and the results of the SYRPINTINE project will be highlighted on the PINDIP website.

One of the projected outcomes of the PINDIP project is the **organization of a symposium** on the 'Role of Diptera in Plant-Pollinator Networks' (see: https://icd9.co.za/symposia-titles/).
at the Ninth International Congress of Dipterology (Windhoek, Namibia, 21-24 November 2018). Evidently, the results of the SYRPINTINE project will be highlighted during the symposium (through a poster and oral communication).

Another projected outcome of the PINDIP project is a **four-monthly *Newsletter***. The issues of the PINDIP *Newsletter* (online only) comprise any aspect on the role of Diptera in pollination biology in Sub-Saharan Africa. The *Newsletter* will highlight the progress and results of the project, other initiatives (e.g. partnerships, congresses and symposia, meeting news, opportunities, etc.), publications, and relevant items that were covered by news media, and with a focus on the Aftrotropics. The *Newsletter* is also published on the PINDIP-website and distributed among a mailing list to which people can subscribe. The results of the SYRPINTINE project will be highlighted in the April 2018 issue of the *Newsletter* (Volume 2, Issue 2).

Training course on the "Taxonomy of pollinating Diptera" (20 November – 1 December 2017): We organized a training course in general Dipterology together with the National Museums of Kenya

and the International Centre of Insect Physiology and Ecology (*icipe*), in Nairobi (Kenya). Fifteen participants of the Afrotropical Regions were educated in general Dipterology, insect collection and insect collection management and the identification of Diptera with an emphasis of pollinating Diptera (Rhiniidae, Calliphoridae and Syrphidae). The results of the SYRPINTINE project were highlighted during the training course.

Kurt Jordaens obtained, together with Olivier Honnay (KULeuven) and Tom Ruttink / Isabel Roldan (ILVO) a FWO-project (FWOG-056517N) entitled: Effects of coffee forest management intensification on arabica coffee (*Coffea arabica* L.) yield, yield quality and wild coffee genetic resources. In this project we will, amongst others, investigate the wild pollinators of arabica coffee in order to estimate the relative contribution of each pollinator species on the pollination of wild arabice coffee.

4. PERSPECTIVES

Besides the projected publications in international journal we will further focus on the whole mtDNA sequencing of the hover flies. We expect the results of the first batch very soon and thereafter we wil produce and sequence a second batch of 60 specimens. Using these data we will be able to resolve the phylogeny of several of the genera of the Afrotropical Syrphidae.

Based upon these results, and results obtained through classical DNA barcoding, we will be able to improve existing, or make new, identification keys for several Afrotropical hover fly genera.

We well also further improve the protocols for the collection of pollen grain in tromical regions and for the sequencing of DNA from pollen grains.

Networking

The research of the SYRPINTINE project has resulted in a strong collaboration between the RMCA and several experts in Syrphidae taxonomy. The ongoing whole mtDNA sequencing is performed in strong collaboration with Dr. Ximo Mengual (Zoologisches Forschungsmuseum Alexander Koenig, Bonn, Germany), Jeffrey H. Skevington (Canadian National Collections of Insects, Arachnids and Nematodes, Toronto, Canada), Alfried Vogler (Imperial College, London, UK), and Xin Zhou (BGI, Beijing, China).

Moreover, the ecological (palynological) research conducted during the SYRPINTINE project has resulted in the following network activities:

- As stated above, in 2017 we organized a first training course on the "Taxonomy of pollinating Diptera" in Kenya (Nairobi). We will organize a second training course on the "Taxonomy of

pollinating Diptera" in the second half of 2019, most probably in South Africa. During these training courses we have/will train enthusiastic African scientist, curators, technicians and students in the taxonomy of Diptera, more specifically those families who have an importance in plant pollination.

- We will organize a symposium on the role of Diptera at the Ninth International Congress of Dipterology in Namibia (Windhoek, 25-30 November 2018). To this end, we will invite and financially support seven participants of the first training course to present their results during the symposium and support the attendants in extending their own research network at the congress.

- Eugène Sinzinkayo was one of the participants of the training course (see above). He recently received a GTI grant of the Royal Belgian Institute of Natural Sciences (RBINS, Brussels) entitled: Study of the impact of habitat degradation on the abundance and diversity of Syrphidae (Insecta: Diptera) in Burundi. In this project, we will, together with Wouter Deconinck (RBINS) develop a first reference collection of the hover flies of Burundi.

Didactic material

The research performed during the SYRPINTINE project on the genus *Eristalinus* will result in a first identification key for the Afrotropical species and will include a new monographic catalogue to the Afrotropical species of the genus. The identification key will contain stacked photographs of the lateral, dorsal and frontal view of most of the species, and photographs of taxonomically relevant characters (e.g. of the abdomen, hind leg, etc).

5. PUBLICATIONS

Apart from the abstract listed above, we intend to publish the results of the SYRPINTINE project as follows:

De Smet, Y., Janssens, S., Ståhls, G., De Meyer, M., Backeljau, T. & Jordaens K. Species delimitation in the hoverfly genus Eristalinus (Diptera, Syrphidae), combining single gene species discovery and coalescent-based species delimitation methods. To be submitted to *PLoS ONE*.

De Smet, Y., Virgilio, M., Sonet, G., Mengual, X., Skevington, J., De Meyer, M., Backeljau, T., Vogler, A., Janssens, S. & Jordaens, K. The full mitochondrial genome of five species of Afrotropical hoverflies (Diptera, Syrphidae, Eristalinus). To be submitted to *mitochondrial DNA*.

Ssymank, A., Jordaens, K., De Meyer, M., Reemer, M. & Rotherary, G. Chapter 60. Syrphidae, In: *The Manual of Afrotropical Diptera*, Volume 3, SANBI, South Africa. The publication of this volume is predicted early 2019.

De Smet, Y., Thompson, F.C., Wyatt, N. & Jordaens, K. A taxonomic revision of the Afrotropical hoverfly genus *Eristalinus* with a new identification key. To be submitted to *ZooKeys*.

6. ACKNOWLEDGEMENTS

Various people have contributed *Eristalinus* specimens for this project. Especially Ximo Mengual (ZFMK-Bonn, Germany), Gunilla Ståhls (MZH-Helsinki, Finland), Jeffrey Skevington (CNC-Toronto, Canada), Axel Ssymank (Germany), Martin Hauser (CAS-San Francisco, USA), Burgert Muller (BMSA-Bloemfontein, South Africa), Ashley H. Kirk-Spriggs (BMSA-Bloemfontein, South Africa), John Midgley (KwaZulu-Natal Museum-Pietermaritzburg, South Africa, Robert Copeland (*icipe*-Nairobi, Kenya) and Laban Njoroge (NMK-Nairobi, Kenya) allowed us to study their material and to take samples for our molecular work.

We especially wish to thank Georg Goergen (IITA-Cotonou, Benin) who not only provided us with numerous Syrphidae specimens from the Afrotropical Region, but also organized the outstanding sampling campaigns in Benin and Togo which allowed us to collect much fresh material needed for the molecular aspects of the SYRPINTINE project, and who helped us with the identification of the plant material.

We wish to thank the members of the Joint Experimental Molecular Unit or the Royal Museum for Central Africa (RMCA – Massimiliano Virgilio) and of the Royal Belgian Institute of Natural Sciences (RBINS – Gontran Sonet, Carl Van Gestel) for their help with the full mtDNA sequencing. Min Tang, Shanlin Xiu and Xin Zhou (BGI-Bejing, China and China Agricultural University) (sequencing), and Alfried Vogler and Hannah Norman (Imperial College-London, UK), Jeffrey Skevington, Andrew Young and Scott Kelso (CNC-Toronto, Canada) (discussion) are acknowledged for the sequencing of the full mtDNA genomes and the various discussions on the improvement of the methodology and interpretation of the data.

Christian F. Thompson provided a first draft of an identification key for the Afrotropical species of the genus *Eristalinus* which was a great help in the identification of the material.

We also wish to thank Wim Baert, Iris Van der Beeten and Pieter Asselman (Botanical Garden Meise) for their help with the molecular analysis of the plant and pollen material and for the SEM of the plant pollen.

This project has also received financial support of the JRS Biodiversity Foundation and the Ministry of Development Cooperation (Belgium)

7. REFERENCES

Allsopp, M.H., de Lange, W.J. & Veldtman, R. (2008). Valuing insect pollination services with cost of replacement. PLoS ONE, 3.

Barrett, C.F. & Freudenstein, J.V. (2011). An integrative approach to delimiting species in a rare but widespread mycoheterotrophic orchid. Molecular Ecology 20, 2771–2786.

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., & Stadler, P.F. (2013). MITOS: improved de novo metazoan mitochondrial genome annotation. Molecular Phylogenetics and Evolution 69, 313-319.

Bezzi, M. (1915). *The Syrphidae of the Ethiopian Region based on material in the collection of the British Museum (Natural History), with descriptions of new genera and species*. British Museum (Natural History), London. 146 pp.

Bickford, D., Bickford, D., Lohman, D.J., Sodhi, N.S., Ng, P.K.L., Meier, R., Winker, K., Ingram, K.K. & Das, I. (2007). Cryptic species as a window on diversity and conservation. Trends in Ecology and Evolution 22, 148–155.

Binladen, J., Gilbert, M.T.P., Bollback, J.P. et al. (2007). The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. PLoS One, 2: e197.

Botero-Castro, F. (2013). Next-generation sequencing and phylogenetic signal of complete mitochondrial genomes for resolving the evolutionary history of leaf- nosed bats (Phyllostomidae). Molecular Phylogenetics and Evolution 69, 728–739.

Bybee, S.M., Bracken-Grissom, H., Haynes, B.D., Hermansen, R., Byers, R.L., Clement, M.J., Udall, J., Wilcox, E.R. & Crandall, K. (2011). Targeted amplicon sequencing (TAS): a scalable next-gen approach to multilocus, multitaxa phylogenetics. Genetics, Biology and Evolution, 3: 1312–23.

Cameron, S.A., Lozier, J.D., Strange, J.P., Koch, J.B., Cordes, N., Solter, L.F. & Griswold, T.L. 2011. Patterns of widespread decline in North American bumble bees. Proc.Natl.Acad.Sci.USA 108: 662-667.

Cameron, S.L. (2013). Insect mitochondrial genomics: Implications for evolution and phylogeny. Annual Review in Entomology 59, 95-117.

Cameron, S.L., Lambkin, C.L., Barker, S.C. & Whiting, M.F. (2007). A mitochondrial genome phylogeny of Diptera: Whole genome sequence data accurately resolve relationships over broad timescales with high precision. Systematic Entomology 32, 40 – 59.

Cameron, S.L., Sullivan, J., Song, H., Miller, K.B. & Whiting, M.F. (2009). A mitochondrial genome phylogeny of the Neuropterida (lace-wings, alderflies and snakeflies) and their relationship to the other holometabolous insect orders. Zoologica Scripta, 38, 575–590.

Cameron, S.L. (2014). How to sequence and annotate insect mitochondrial genomes for systematic and comparative genomics research. Systematic Entomology 39, 400–411.

Carstens, B.C., Pelletier, T.A., Reid, N.M. (2013). How to fail at species delimitation. Molecular Ecology 22, 4369–4383.

Choi, S.C. & Hey, J. (2011). Joint inference of population assignment and demographic history. Genetics, 189, 561–577.

Curran, C.H. (1939). Records and descriptions of African Syrphidae.–IV (Diptera). American Museum Novitates, 1026, 10 pp.

Darriba, D., Taboada, G.L., Doallo, R. & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. Nature Methods 9, 772.

De Smet, Y., De Clerck, O., Uemachi, T., Granados Mendoza, C., Wanke, S., Goetghebeur, P. & Samain, M.-S. (2017). Multilocus coalescent species delimitation to evaluate traditionally defined morphotypes in *Hydrangea* sect. *Asperae* (Hydrangeaceae). Molecular Phylogenetics and Evolution 114, 415-425.

Dessein, S., Harwood, R., Smets, E. & Robbrecht, E. (2005). Pollen of the Spermacoce (Rubiaceae) species from the Northern Territory of Australia: morphology and taxonomic significance. Australian Systematic Botany 18: 367-382.

De Queiroz, K. (2011). Branches in the lines of descent: Charles Darwin and the evolution of the species concept. Biological Journal of the Linnean Society 103, 19–35.

De Queiroz, K. (2007). Species concepts and species delimitation. Systematic Biology 56, 879–886.

De Queiroz, K. (1999). The general lineage concept of species and the defining properties of the species category, in: *Species: New Interdisciplinary Essays*. pp. 49–89.

De Queiroz, K. (1998). Chapter 5. The general lineage concept of species, species criteria, and the process of speciation: A conceptual unification and terminological recommendations., in: *Endless Forms: Species and Speciation*. pp. 57–75.

Dirickx, H.G. (1998). *Catalogue Synonymique et Géographique des Syrphidae (Diptera) de la Région Afrotropicale*. Museum d'Histoire Naturelle, Genéve, 187 pp.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32, 1792-1797.

Ence, D.D. & Carstens, B.C. (2010). SpedeSTEM : a rapid and accurate method for species delimitation. Molecular Ecology Resources 11, 473–480.

Fischer, O. A., Mátlová, L., Dvorská, L., Svástová, P., Bartos, M., Weston, R.T. & Pavlík, I. (2006). Various stages in the life cycle of syrphid flies (*Eristalis tenax*; diptera: Syrphidae) as potential mechanical vectors of pathogens causing mycobacterial infections in pig herds. Folia Microbiologica 51, 147–153.

Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from metazoan invertebrates. Molecular Marine Biology and Biotechnology 3, 294-299.

Fujisawa, T. & Timothy, G.B. (2013). Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent approach : A revised method and evaluation on simulated data sets. Systematic Biology 62, 707–724.

Gibson, J.F., Kelso, S., Jackson, M.D., Kits, J.H., Miranda, G.F.G. & Skevington, J.H. (2011). Diptera-specific polymerase chain reaction amplification primers of use in molecular phylogenetic research. Annals of the Entomological Society of America 104, 976-997.

Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E.M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S. & Russ, C. (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nature Biotechnology 27, 182-189.

Grosskopf, G. (2005). Biology and life history of *Cheilosia urbana* (Meigen) and *Cheilosia psilophthalma* (Becker), two sympatric hoverflies approved for the biological control of hawkweeds (*Hieracium* spp.). New Zealand Biological Control 35, 142–154.

Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H. & Hallwachs, W. (2004). Ten species in one : DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proceedings of the National Academy of Sciences of the USA 101, 14812–14817.

Heled, J. & Drummond, A.J. (2010). Bayesian inference of species trees from multilocus data. Molecular Biology and Evolution 27, 570–580.

Huw A.O., Heled, J., Xie, D. & Drummond, A.J. (2016). Computational performance and statistical accuracy of *BEAST and comparisons with other methods. Systematic Biology 65, 381–96.

Janssens, S.B., Song Wilson, Y., Yuan, Y.-M., Nagels, A., Smets, E.F. & Huysmans, S. 2012. A total evidence approach using palynological characters to infer the complex evolutionary history of the Asian *Impatiens* (Balsaminaceae). Taxon 61, 392-404.

Junqueira, A. C., Azeredo-Espin, A.M., Paulo, D.F., Marinho, M.A., Tomsho, L.P., Drautz-Moses, D.I., Purbojati, R.W., Ratan, A. & Schuster, S.C. (2016). Large-scale mitogenomics enables insights into Schizophora (Diptera) radiation and population diversity. Scientific Reports, 6.

Klein, A.M., Vaissière, B., Cane, J.H., Steffan-Dewenter, I., Cunningham, S.A., Kremen, C. & Tscharntke, T. 2007. Importance of crop pollinators in changing landscapes for world crops. Proceedings of the Royal Society of London B, Biological Sciences 274, 303-313.

Klein, A.M., Mueller, C.M., Hoehn, P. & Kremen, C. 2009. Understanding the role of species richness for pollination services. In: Bunker, D., Hector, A., Loreau, M., Perrings, C. & Naeem, S. (eds.): Biodiversity, Ecosystem Functioning, and Human Wellbeing: An Ecological and Economic Perspective. Oxford University Press, NY, pp. 195-208.

Kim, S.-T., Sultan, S.E. & Donoghue, M.J. (2008). Allopolyploid speciation in *Persicaria* (Polygonaceae): insights from a low-copy nuclear region. Proceedings of the National Academy of Sciences of the U.S.A. 105, 12370–12375.

Kremen, C. 2008. Crop pollination services from wild bees. In: James, R. & Pitts-Singer, T.L. (eds.): Bee Pollinators in Agricultural Ecosystems. Oxford University Press, NY, pp. 10-26.

Kremen, C., Williams, N.M. & Thorp, R.W. 2002. Crop pollination from native bees at risk from agricultural intensification. Proc.Natl.Acad.Sci.USA 99: 16812-16816.

Kuchta, S.R., Brown, A.D., Converse, P.E. & Highton, R. (2016). Multilocus phylogeography and species delimitation in the Cumberland Plateau salamander, *Plethodon kentucki*: Incongruence among data sets and methods. PLoS One 1–25.

Leaché, A.D. & Fujita, M.K. (2010). Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). Proceedings of the Royal Society of London, Biological Sciences 277, 3071–3077. doi:10.1098/rspb.2010.0662

Lemmon, A.R., Brown, J.M., Stanger-Hall, K., Lemmon, E.M. (2009). The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference. Systematic Biology 58, 130–145.

Li, X., Ding, S., Li, X., Peng, H., Tang, C. & Yang, D. (2017). The complete mitochondrial genome analysis of *Eristalis tenax* (Diptera, Syrphidae). Mitochondrial DNA 2, 654-655.

Li, X., Lin, L., Cui, A., Bai, J., Wang, X., Xin, C., Zhang, Z., Yang, C., Gao, R., Huang, Y. & Lei, F. (2016). Taxonomic status and phylogenetic relationship of tits based on mitogenomes and nuclear segments. Molecular Phylogenetics and Evolution 104, 14–20.

Logue, K., Chan, E.R., Phipps, R., Small, S., Reimer, L., Henry-Halldi, C., Sattabongkot, J., Siba, P.M., Zimmerman, P.A. & Serre, D. (2013). Mitochondrial genome sequences reveal deep divergences among *Anopheles punctulatus* sibling species in Papua New Guinea. Malaria Journal 12, 64.

Ma, C., Yang, P.C., Jiang, F., Chapuis, M.-P., Shall, Y., Sword, G.A. & Kang, L. (2012). Mitochondrial genomes reveal the global phylogeography and dispersal routes of the migratory locust. Molecular Ecology 21, 4344–4358.

Mik, J. (1897). Einige Bemerkungen zur Dipteren-Familie der Syrphiden. Wiener Entomologische Zeitung 16, 61-66.

Miller, M.A., Pfeiffer, W. & Schwartz, T. (2010). Creating the CIPRES science gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA pp 1-8.

Müller, K., Müller, J., Neinhuis, C. & Quandt, D. (2008) PhyDE – Phylogenetic data editor, Version 0.9971. http://www.phyde.de/.

Nelson, L.A., Lambkin, C.L., Batterham, P., Wallman, J.F., Dowton, M., Whiting, M.F., Yeate, D.K. & Cameron, S.L. (2012). Beyond barcoding: genomic approaches to molecular diagnostics in blowflies (Diptera: Calliphoridae). Gene 511, 131–142.

Pérez-Bañón, C., Rojo, S., Ståhls, G. & Marcos-García, M.A. (2003). Taxonomy of European *Eristalinus* (Diptera : Syrphidae) based on larval morphology and molecular data 100, 417–428.

Pimentel, R. & Riggins, R. 1987. The nature of cladistic data. Cladistics 3, 201–209.

Pogue, M & Mickevich, M. (1990). Character definitions and character-state delimitations: The bête noire of phylogenetic inference. Cladistics 6, 319–361.

Pons, J., Barraclough, T.G., Urita, J.E.G.O., Ardoso, A.N.C., Uran, D.A.P.D., Azell, S.T.H., Amoun, S.O.K., Umlin, W.I.D.S. & Ogler, A.L.P.V. (2006). Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Systematic Biology 55, 595–609.

Pramual, P. & Kuvangkadilok, C. (2012). Integrated cytogenetic , ecological , and DNA barcode study reveals cryptic diversity in *Simulium* (*Gomphostilbia*) *angulistylum* (Diptera: Simuliidae). Genome 12, 1–12.

Psonis, N., Antoniou, A., Kukushkin, O., Jablonski, D., Petrov, B., Crnobrnja-Isailovic, J., Sotiropoulos K., Gherghel I., Lymberakis P. & Poulakakis, N. (2017). Hidden diversity in the *Podarcis tauricus* (Sauria, Lacertidae) species subgroup in the light of multilocus phylogeny and species delimitation. Molecular Phylogenetics and Evolution 106, 6–17.

Pu, D.-Q, Liu, H.-L., Gong, Y.-Y., Ji, P.-C., Li, Y.-J., Mou, F.-S & Wei, S.-J. (2017). Mitochondrial genomes of the hoverflies *Episyrphus balteatus* and *Eupeodes corollae* (Diptera : Syrphidae), with a phylogenetic analysis of Muscomorpha. Scientific Reports 7, 44300.

Puillandre, N., Lambert, A., Brouillet, S. & Achaz, G. (2011). ABGD, Automatic Barcode Gap Discovery for primary species delimitation. Molecular Ecology 8, 1864-1877.

Ramaiya, M., Johnson, M.G., Shaw, B., Heinrichs, J., Hentschel, J., von Konrat, M., Davison, P.G. & Shaw, A.J. (2010). Morphologically cryptic biological species within the liverwort *Frullania asagrayana*. American Journal of Botany 97, 1707–1718.

Rambaut, A. & Drummond, A.J. (2013). Tracer V1.6. Available from http//beast.bio.ed.ac.uk/Tracer.

Raupach, M.J., Amann, R., Wheeler, Q.D. & Roos, C. (2016). The application of "-omics" technologies for the classification and identification of animals. Organisms Diversity & Evolution 16, 1-12.

Reeves, A.P. & Richards, M.C. (2011). Species delimitation under the General Lineage Concept : An empirical example using wild North American hops (Cannabaceae : *Humulus lupulus*). Systematic Biology 60, 45–59.

Rondani, C. (1845). Ordinamento sistematico dei generi italiani degli insetti ditteri. Nuovi Annali Delle Scienze Naturali 1884, 443–459.

Setiadi, M.I., McGuire, J.A., Brown, R.M., Zubairi, M, Iskandar, D.T, Andayani, N., Supriatna, J. & Evans, B.J. (2011). Adaptive radiation and ecological opportunity in Sulawesi and Philippine Fanged Frog (*Limnonectes*) communities. American Naturalist 178, 221–240.

Shaw, K.L. (2002). Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: What mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. Proceedings of the National Academy of Sciences of the USA 99, 16122-16127.

Shiraki, T. (1930). Die Syrphiden des Japanischen Kaiserreichs, mit Berrücksichtigung benachbarter Gebiete. Memoirs of the Faculty of Agriculture, Taihoku Imperial University 1, 1-446.

Sites, J.W. & Marshall, J.C. (2003). Delimiting species : a Renaissance issue in systematic biology. Trends in Ecology and Evolution 18, 462–470.

Sites, J.W. & Marshall, J.C. (2011). Operational criteria for delimiting species. Annual Review in Ecology, Evolution and Systematics 35, 199–227.

Ståhls, G., Hippa, H., Rotheray, G., Muona, Jj. & Gilbert, F. (2003). Phylogeny of Syrphidae (Diptera) inferred from combined analysis of molecular and morphological characters. Systematic Entomology 28, 433–450.

Ståhls, G., Vujic, A., Pérez-Bañon, C., Radenkovic, S., Rojo, S. & Petanidou, T. (2009). COI barcodes for identification of *Merodon* hoverflies (Diptera, Syrphidae) of Lesvos Island, Greece. Molecular Ecology Resources 9, 1431-1438.

Stevens, P.F. (1991). Character states, morphological variation, and phylogenetic analysis: A review. Systematic Botany 16, 553–583.

Smith, K. G. V. & Vockeroth, J. R. (1980). Family Syrphidae. Pp. 488-510. In: Crosskey, R. W. (ed.), *Catalogue of the Diptera of the Afrotropical Region*. 1437 pp. British Museum (Natural History), London. [1980.07.10]

Tang, M., Tan, M., Meng, G., Yang, S., Su, X., Liu, S., Song, W., Li, Y., Wu, Q., Zhang, A. & Zhou, X. (2014). Multiplex sequencing of pooled mitochondrial genomes – a crucial step toward biodiversity analysis using mito-genomics. Nucleic Acids Research 42, e166-e166.

Thompson, F.C. (2003). *Austalis*, a new genus of flower flies (Diptera: Syrphidae) with revisionary notes on related genera. Zootaxa 246, 1–19.

Thompson, F.C. & Rotheray, G. (1998). Family Syrphidae. In: Papp, L. & Darvas, B. (eds.): *Contributions to a Manual of Palaearctic Diptera (with Special Reference to Flies of Economic Importance)* Vol. 3, Higher Brachycera. Science Herald, Budapest, 81-139.

Van de Putte, K., Nuytinck, J., De Crop, E. & Verbeken, A. (2016). *Lactifluus volemus* in Europe: Three species in one - Revealed by a multilocus genealogical approach, Bayesian species delimitation and morphology. Fungal Biology 120, 1–25.

van Engelsdorp, D., Evans, J.D., Saegerman, C., Mullin, C., Haubruge, E., Nguyen, B.K., Frazier, M., Frazier, J., Cox-Foster, D., Chen, Y.P., Underwood, R., Tarpy, D.R. & Pettis, J.S. (2009). Colone Collapse Disorder: a descriptive study. PLoS ONE, 4.

Verlinden, L. (1994). Faune de Belgique. Syrphides (Syrphidae). Institut Royal des Sciences Naturelles de Belgique, Bruxelles. 289 pp.

Welton, L.J., Siler, C.D., Oaks, J.R., Diesmos, A.C. & Brown, R.M. (2013). Multilocus phylogeny and Bayesian estimates of species boundaries reveal hidden evolutionary relationships and cryptic diversity in Southeast Asian monitor lizards. Molecular Ecology 22, 3495–3510.

White, T.R., Conrad, M.M., Tseng, R., Balayan, S., Golding, R., Manuel, A., Martins, D.F. & Dayrat, B.A. (2011). Ten new complete mitochondrial genomes of pulmonates (Mollusca : Gastropoda) and their impact on phylogenetic relationships. BMC Evolutionary Biology 11, 295.

Whittington, A.E. (2003). The Afrotropical Syrphidae fauna: an assessment. Studia dipterologica 10, 579-607.

Wiens, J. J. (1995). Polymorphic characters in phylogenetic systematics. Systematic Biology 44, 482–500.

Wiens, J.J. & Penkrot, T.A. (2002). Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (*Sceloporus*). Systematic Biology 51, 69–91.

Winfree, R., Williams, N.M., Dushoff, J. & Kremen C. (2007). Wild bees provide insurance against ongoing honey bee losses. Ecology Letters 10, 1105-1113.

Wu, L.-W., Lin L.-H., Lees, D.C. & Hsu, Y.-F. (2014). Mitogenomic sequences effectively recover relationships within brush-footed butterflies (Lepidoptera: Nymphalidae). BMC Genomics 15, 468.

Yang, Z. & Rannala, B. (2010). Bayesian species delimitation using multilocus sequence data. Proceedings of the National Academy of Sciences of the USA 107, 9264–9269.

Yong, H.-S., Song, S.-L., Lim, P.-E., Chan, K.-G., Chow, W.-L. & Eamsobhana, P. (2015). Complete mitochondrial genome of *Bactrocera arecae* (Insecta: Tephritidae) by next generation sequencing and molecular phylogeny of Dacini tribe. Scientific Reports 5.

Zhang, C., Zhang, D.-X., Zhu, T. & Yang, Z. (2011). Evaluation of a bayesian coalescent method of species delimitation. Systematic Biology 60, 747-761.

Zhang, J., Kapli, P., Pavlidis, P. & Stamatakis, A. (2013). A general species delimitation method with applications to phylogenetic placements. Bioinformatics 29, 2869–2876.

Zhao, Z., Su, T., Chesters, D., Wang, S., Ho, S.Y.W., Zhu, C. & Chen, X. (2013). The mitochondrial genome of *Elodia flavipalpis* Aldrich (Diptera: Tachinidae) and the evolutionary timescale of tachinid flies. PLoS ONE 8, e61814.

Zhou, X., Li, Y., Liu, S., Yang, Q., Su, X., Zhou, L., Tang, M., Fu, R., Li, J. & Huang, Q. (2013). Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplication. Gigascience 2, 4.