Morphological variation and strain identification of insects using wings and I^3S.

Nayna Vyas-Patel[1], John D Mumford[2].

[1] Runnemede Bioscience, Surrey, UK. Runnemede.Bioscience@gmail.com.

[2] Imperial College, London, UK. j.mumford@imperial.ac.uk

Key Words: Diptera:Culicidae, Mosquito Species and Strains, Image Recognition Software, I^3S Classic, CO1, *Anopheles gambiae, Anopheles arabiensis, Anopheles coluzzii, Anopheles stephensi, Aedes aegypti, Aedes albopictus, Coquillettidia perturbans, Culex quinquefasciatus, Ochlerotatus taeniorhynchus, Ochlerotatus triseriatus, Ochlerotatus sollicitans, Psorofora ferox,* transgenic, Insect Wings.

Abstract

Novel insect identification techniques often lead to speculation on whether the method could cope with any intraspecific variation that might occur in a species. Using I^3S Classic (Interactive Individual Identification System, Classic) and images of mosquito wings, different mosquito strains were tested with a copy of the strain present or absent from the database which contained images of other strains of the test species. When a wing image of the exact species, strain and sex was present in the database, there was 100% (or near 100%) retrieval of the correct species and strain at rank one. When the exact strain was absent from the database, but other strains of the same species were present, the retrieval rates at rank one were again high (100%) in the majority of cases and when they were not, the correct species was generally retrieved at rank two. Out of 40 different species and strains tested, only three were significantly different at rank one when the exact strain was absent from the database. In general, images of field strains selected for each other and therefore were similar to each other in greater numbers and instances than for the laboratory strains tested. When a copy of a strain was absent from the database, but other strains/sibling species were present, I^3S retrieved the correct strains/sibling species at rank one in the majority of cases. In the one case of transgenic mosquitoes tested, I^3S could reliably be used to identify transgenic mosquitoes from the parent stock as they were retrieved 100% at rank one when both the transgenic

and unmodified parent strains were present in the database. This indicates the potential of using I^3S to distinguish transgenic or other selectively bred strains from a parent strain, also selectively bred and wild mosquitoes, at least in the first phase after field release. Similarly, hybrid strains, sibling species and members of species complexes as in the *Anopheles gambiae* species complex could also be correctly identified when copies of all the relevant species/strains/siblings were in the database. This contradicts the belief that only molecular characterisation could separate *A. gambiae* s.s. from *A. coluzzii,* or *A. arabiensis;* I^3S could accurately separate them all. I^3S worked as it was set up to do, retrieving closely resembling images of the test insects from the database and ranking them in order of similarity. Dealing with any intraspecific variation was therefore not an issue if the software (I^3S) was used systematically. I^3S complements molecular and traditional taxonomic methods for species identification and the separation of sibling complexes and strains. In future, it should become the norm to maintain databases of insect wings and other body part images for use in image recognition.

Introduction

Morphological variation, large or small, exists in many organisms including insects and could potentially give rise to challenges in accurate species identification (Keeley 1982; Nosil & Reimchen, 2005; Ampuero et al 2009 & 2010 and Paz Garcia et al, 2015). This variation in appearance and size can occur over distances when a given species might present with differences in phenotype and/or size and is a feature of many island species that look different from their siblings on the mainland; or in the same species that live many miles or continents apart. It also occurs in species that experience different environmental conditions to their siblings living elsewhere. Termed ‘ecological variation’ it has been documented in many organisms including insects (Yi Bai et al 2016, Suman et al 2009, Dellicour et al 2017, Ekgachai et al 2013).

The wings of insects have long been known to be a reliable diagnostic feature for a given species (Woodward 1926, Wooton 1992, Wilke et al 2016). The use of algorithms and software to capture wing shape and the pattern and proportions of the vein structure on insect wings to differentiate species, means that even the

smallest change in this shape, pattern and/or proportion can be detected quickly and easily. Image recognition software such as I^3S Classic (the open source, Interactive Individual Identification System, I^3S) can detect those changes far better than the human eye. Even if a test specimen was smaller or larger than the correct species present in the database, I^3S takes account of the pattern and geometry of the veins on the test specimen to retrieve the closest reference image of a species, if not at rank one then at least within the first five ranks (Vyas-Patel et al 2016; Vyas-Patel & Mumford, 2017). Despite this use of mathematics, geometry and software, one of the criticisms for the use of technology and other 'automated identification systems' in insect species identification, has been the unverified argument (for I^3S at least) that different strains of a given species might vary (intraspecific variation) leading to inaccuracies in species identification. Jean-Pierre Dujardin explained at great length why modern morphometric methods including image recognition, are capable of detecting even the smallest of change in phenotypic features that can exist between morphologically similar sibling species and aid the accurate identification of species and strain recognition (Dujardin, 2011).

The concept 'different strains of a species' is understood well when referring to the genotype where a 'strain' is a genetic variant or subtype of an organism. Different 'strains', 'sibling species' and 'sub species' are generally considered to be genetic variants of a given species. Variation can also occur in the phenotype and it was this 'appearance based', intraspecific variation within species that was of interest in this study. This can be assessed using geometric morphometric techniques or its modern equivalent - image recognition software. Insects respond to stress and conversely the lack of stress (very favourable conditions) in many different ways for example by changing their behaviour or by changes in the pattern of their life histories. The response may also result in changes in their morphology and physiology. It is increasingly being realized that it is not only distances or geographical isolation that can lead to phenotypic variation, but also factors such as nutrition (Pieterse et al 2017) and soil type (Benitez et al 2013). Pieterse et al (2017) noted that fruit flies reared on different fruits, notably apple and pear, had significantly different wing shapes compared to fruit flies reared on other fruits and that this could be detected and measured using geometric morphometric techniques to gauge wing shapes. Benitez (2013) demonstrated that using geometric morphometrics, the shape of the

hind wings of *Diabrotica virgifera virgifera* changed according to the major soil types of Croatia. The wing morphology of two sibling *Drosophila* species was influenced by the different host plants (cactus) they fed on (Carreira et al 2006; Soto et al 2008 & 2010). Gomez-Cendra et al (2016) found that it was not geographical distance that affected the wing morphology of their fruit flies, but the host fruit available to the flies, those feeding on the same fruit had similar wing morphology no matter how far apart they lived. Using molecular markers and noting phenotypic differentiation, Gomez-Cendra et al (2016) reported that flies collected from and feeding on peaches or walnuts differed genetically and in appearance from each other regardless of geographical or temporal overlap. Yi Bai found that the grasshopper species *Trilophidia annulata* followed the Bergmann rule where individuals from higher, cooler latitudes with longer growing seasons were larger in size (using wing length as a measure) than individuals from the lower, warmer latitudes (Yi Bai et al 2016). There could be many other unknown factors that could potentially have an effect on phenotypic variation. Therefore, the present study considered any species collected from different locations as a 'strain', this included species reared in different laboratories and separate colonies within laboratories; if the same species was reared in different laboratories, they were considered as different strains of the species for the purposes of this study. Some of the species reared in laboratories were hybrids, created to prevent inbreeding, others were reared because they were known to be strains which were either resistant or susceptible to insecticides and/or susceptible or refractory to medically important parasites or viruses such as the malaria parasite or the Chikungunya, Zika and Dengue viruses. Both field caught and laboratory reared species and strains were examined. Mosquitoes were used as the test insect.

As evidenced above, a large body of information exists on the use of geometric morphometric methods to distinguish between insect species using wings. He-Ping Yang's (2015) account of different geometric morphometric methods to identify insects using wing images currently provides the most encompassing review. Wilke et al (2016) dealt entirely with methods using wing morphometrics for the identification of different mosquito species. This comprehensive study is amongst the few to describe the use of software and image recognition for the identification of species and strains. The project used the freely downloadable software, I^3S Classic

to test the identification of different strains of insect species (both laboratory reared and field caught) to determine if it could retrieve the correct species and strain when a reference image of the test taxon was not in the database but other strains of the test species were.

Ordinarily, as I^3S is very accurate at retrieving an image of the test species if it is present in the database (Vyas-Patel & Mumford 2017), there would be no point in trying to retrieve a copy of an image which was not present in the database. In such cases, only the closest match could be retrieved and the score attached to each ranked image in the results would give some indication as to how far the match at rank 1 was from the test. Knowledge of the contents of the database was emphasised in previous studies using I^3S, as the software cannot fully match an image of a species which was not present in the database (Vyas-Patel & Mumford 2017); only the closest match could be retrieved. However, for the purpose of this study, example images of all the species and strains to be tested were first uploaded into the database (one wing image of each species, strain and sex) and each species and strain was tested. Next, the test strain reference was removed from the database, leaving other, different strain/s of the same species in the database and re-tested. Hence strains were tested with and then without a copy of the specific reference strain in the database.

Most image recognition software is designed to compare like for like (similarities) and rank the results. Hence, even if a specific reference image was present in the database, if the test had not been aligned in the same way as the database image of the same species, or marked in the same way, it could result in errors. Care was taken to align and mark the wings consistently, in the same way for both the test and database images. Alignment was achieved by eye using software to rotate the image by degrees (custom rotate, Photoshop), pin point precision was not required. The wing samples used were largely intact (unbroken) except in the few cases of field caught specimens where the numbers available for testing were low. In such cases, imperfect, broken wings were also imaged and tested. Any major folds if present were gently eased out by the placement of a coverslip over the wing before a photograph was taken. This was particularly important for smaller wings such as in mosquitoes, which are generally in the order of 3mm long, but less so for larger

insect wings, where the stronger vein structure tended to keep the wings flat (experience from previous studies using Hymenopteran wings).

Materials and Method

Insect Specimens

Mosquito strains originating from around the world and reared in different laboratories, as well as field caught strains from different locations were used for the study. The species were identified by trained scientists in every case. The donors and the available metadata for each strain are given in Table 1 together with footnotes.

Preparation and marking of the Images.

The wings were dissected from the body under a standard dissection microscope, photographed with a Samsung NV10 digital camera, using only the sub-stage lighting of the microscope, as this produced a clear image of the wing shape and venation. Prior to taking a photograph, a clean microscope cover slip was placed on the wing to ensure that any folds were gently flattened out. Each image was uploaded into an Adobe Photoshop (CS6) image editor and rotated so that the point of insertion of the wing into the body of the insect always faced to the left and the wing was aligned to be as horizontal as possible, using 'Image rotate' and 'Custom rotate' in the top menu bar of Photoshop. The newly aligned and rotated images were saved as .jpg files, creating a different file for each species and sex.

I^3S Classic, version 4.02, was downloaded from the internet and a 'fingerprint' (.fgp) image of each wing image was prepared and stored in the database as described in previous studies (Vyas-Patel & Mumford 2017) and shown in Figure 1. Any clearly visible point and landmark wing feature may be chosen for marking as long as subsequent images were also consistently marked in the same way. Each fingerprint file was saved and could then be used either for reference database creation or as a test image. A comprehensive guide to the use of I^3S and how images can be prepared, stored and analysed is given in the instructions on the I^3S website together with a tutorial (http://www.reijns.com/i3s/download/I3S%20Classic.pdf).

Database creation.

A database of fingerprint (.fgp) files was created. This comprised one image of each sex and strain of *Aedes aegypti* (both field caught and laboratory reared), *Aedes albopictus* (field caught and laboratory reared), *Aedes vexans* (field caught), *Anopheles arabiensis* (laboratory reared), Transgenic Cas9 derived from hybrid *Anopheles gambiae* G3 reared at Imperial College London (ICL), *Anopheles gambiae* G3 (ICL) *Anopheles gambiae* (Kisumu & Tiassale strains), *Anopheles gambiae N'guesso (A. coluzzii*), *Anopheles gambiae Keele (A. coluzzii*), *Anopheles stephensi* (laboratory reared, strains Red, Beech, SDA500, STML and Dub5), *Coquillettidia perturbans* (field caught), *Culex quinquefasciatus* (field caught and laboratory reared - Muheza, TPRI, UG), *Ochlerotatus taeniorinchus* (field caught), *Ochlerotatus triseriatus* (field caught), *Ochlerotatus sollicitans* (field caught), and *Psorophora ferox* (field caught). A total of forty different strains and sixty different wing images of mosquito strains were uploaded into the database. Wing images of both sexes were included from the laboratory reared species/strains. In the case of field caught specimens only females were available, as the traps were designed to attract adult female mosquitoes and so largely collected females. Furthermore, in the case of field caught specimens where the numbers available were low, broken edged (imperfect) wings were also used. Images were added and removed from the database as required for each of the comparisons. The test images were therefore assessed with and without a corresponding reference wing image in the database and rank 1 and 2 results were noted in each case, Table 2.

Any significant difference in results when the test image was present compared to absent from the database is given in Table 3. The salient features of the observed results for each strain is given in Table 4.

Results

Table 1. Strains and species of Mosquitoes used with available metadata.

Aedes aegypti IAEA	Originating from Brazil, laboratory reared at the International Atomic Energy Agency (IAEA), Seibersdorf, Austria.
Aedes aegypti	From West Africa, insecticide susceptible, refractory to

LSHTM	Filarial infections. Laboratory reared since 1926, London.
Aedes aegypti UG	Field caught, University of Georgia, USA.
Aedes albopictus UG	Field caught from Fayetteville, Alabama, by University of Georgia, USA.
Aedes albopictus SAV	Field caught from Savannah, Georgia, USA.
Aedes albopictus IAEA	Originating from Rimini, Italy and laboratory reared at IAEA.
Aedes vexans UG	Field caught from Clinton, South Carolina, Laurens County, Ex River Bottom, by University of Georgia, USA.
Aedes vexans Den	Field caught from Denver, USA, +40.18920N, -105.05975W, Alt. 4,995 ft.
Aedes vexans	Field caught from the Atlanta, Georgia region, USA.
Anopheles arabiensis IAEA	Originating from Dongola, Sudan, Africa. Laboratory reared at IAEA and used in the Sterile Insect Technique, SIT.
Anopheles arabiensis ICL	Obtained from MR4, BEI resources (Malaria Research & Reference Reagent Resource Centre), the strain was laboratory reared at Imperial College London (ICL) and originates from Dongola, Sudan.
Anopheles gambiae G3, ICL	Laboratory reared at Imperial College London (ICL), UK, the strain G3 ICL, was obtained from MR4, BEI resources. Originally, the strain was collected from McCarthy Island, The Gambia, West Africa, but is now known to be a hybrid.
Anopheles gambiae G3, LSHTM	Laboratory reared at the London School of Hygiene & Tropical Medicine (LSHTM) since 1998 for its susceptibility to Dieldrin.
Transgenic Cas9 (from *Anopheles gambiae G3)*	The Cas9 transgenic mosquitoes were genetically modified using the G3 strain at Imperial College London (ICL) and selected for fluorescent marker expression at each generation, (Hammond et al 2016).
Anopheles gambiae	Laboratory reared at LSHTM from 1998 onwards, from

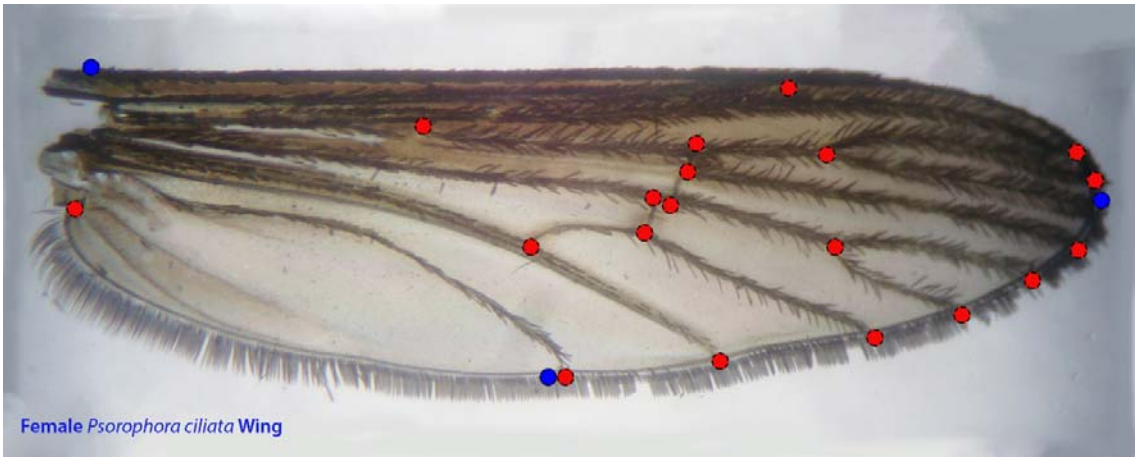
Kisumu	Kisumu, Kenya, Africa. Fully susceptible to Pyrethroids.
Anopheles gambiae Tia (Tiassale)	Laboratory reared at LSHTM from 2013 onwards. From Tiassale, Cote d'Ivoire, Africa.
Anopheles gambiae N'guesso (A. coluzzii) ICL	Colonised in 2006 at ICL, from mosquitoes collected around Yaoundé Cameroon and reclassified as *A. coluzzii* following molecular characterisation, Habtewold et al, 2016.
Anopheles gambiae N'guesso (A. coluzzii) LSHTM	Laboratory reared at LSHTM since 2013, this strain was obtained from ICL (not known at the time of testing).
Anopheles gambiae Keele (A. coluzzii)	Laboratory reared at the University of Glasgow. Originally a hybrid colony from crosses between four different *A. gambiae* colonies maintained at Keele University; molecular characterisation indicated that the Keele colony was *A. coluzzii* (Ranford-Cartwright, 2016).
Anopheles stephensi SDA500	Laboratory reared at LSHTM from a mixture of stocks from Sind province, Pakistan.
Anopheles stephensi Red	Maintained at LSHTM, from a colony with genetically recessive red eyes.
Anopheles stephensi Beech	From New Delhi India, colonised since 1947 at Beecham's laboratory and subsequently reared at the LSHTM.
Anopheles stephensi STML	Laboratory reared at LSHTM. From Lahore, Pakistan, colonised since 1979.
Anopheles stephensi Dub5	Laboratory reared at LSHTM. From Dubai, colonised since 1989.
Coquillettidia perturbans UG	Field collected by University of Georgia (UG) entomologists from the Georgia area, USA.
Coquillettidia perturbans Sav	Field collected by the Centre for Disease Control (CDC) entomologists, from Savannah, Georgia, USA.
Coquillettidia perturbans Den	Field collected from Denver, +39.93760N,-105.05740W, Alt 5,380 ft.
Coquillettidia	Field collected from the Atlanta region (Atl), Georgia,

perturbans Atl	USA.
Culex quinquefasciatus Mu	Laboratory reared at the LSHTM, pyrethroid resistant. From Muheza (Mu) Tanzania.
Culex quinquefasciatus TPRI	From a colony maintained at the Tropical Pesticide Research Institute (TPRI), Tanzania. Reared at LSHTM since 2010.
Culex quinquefasciatus Sav	Field collected from the Savannah (Sav) region, USA.
Culex quinquefasciatus UG	Field collected from Clarke County by University of Georgia (UG) Athens, Georgia, USA.
Ochlerotatus taeniorinchus Sav	Field collected from the Savannah, Georgia region, USA.
Ochlerotatus taeniorinchus UG	Field collected from Darien, McIntosh County, Georgia, by University of Georgia entomologists, USA.
Ochlerotatus triseriatus Sav	Field collected from the Savannah, Georgia region, USA.
Ochlerotatus triseriatus UG	Field collected from Clarke County, Georgia, by University of Georgia entomologists, USA.
Ochlerotatus sollicitans Sav	Field collected from the Savannah, Georgia region, USA.
Ochlerotatus sollicitans UG	Field collected from Darien, McIntosh County, Georgia, by University of Georgia entomologists, USA.
Psorophora ferox Sav	Field collected from the Savannah, Georgia region, USA.
Psorophora ferox UG	Field collected by University of Georgia, Athens, Georgia, USA.

The different mosquito species were identified by the donors, all of them experienced mosquito taxonomists, using traditional mosquito identification keys.

Mosquito donors: Centre for Disease Control (CDC) Atlanta (Atl), donor: Dr Rosmarie Kelly; University of Georgia (UG), donor: Dr Elmer Gray; Colorado Mosquito Control CDC, Denver (Den), donor: Dr Michael Weissman; International Atomic Energy Agency IAEA, donor: Dr Jorge Hendrichs; Chatham County Mosquito Control, Savannah (Sav), donor: Dr Laura Peaty & Dr Robert Moulis; Imperial College London, ICL donor: Dr Nikolai Windbichler; London School of Hygiene & Tropical Medicine (LSHTM), donor: Miss Shahida Begum and Prof Mark Rowland. University of Glasgow donor: Dr Lisa Ranford-Cartwright.

Figure 1: A mosquito wing indicating the markings made using I^3S.



The female *Psorophora ciliata* mosquito above, was identified and donated by Dr Elmer Gray, University of Georgia Athens, USA and was field caught in Georgia. The red and blue markings have been exaggerated for clarity. The wings were aligned to be as horizontal as possible with the curved part of the wing to the right hand side and the point of insertion of the wing into the insect thorax to the left; before marking in I^3S. The blue dots are used by the software for global alignment when comparing two images and were marked first (as dictated by the software). In this study the first blue dot was made at the top of the wing, in line with the kink where the bottom of the wing curves to meet the point of insertion of the wing into the thorax. The second blue dot was made on the far right end of the wing (at the longest point on the curve of the wing). The third is at the bottom of the wing (widest part), so all of them outline the edge of the wing. The red dots were made at points where the veins meet the periphery of the wing and also their junctions within the wing.

Table 2. Total percentage of strains correctly identified at Ranks 1 & 2, with and without an image of the test strain in the database (DB).

Test Strain	R1 Test **in** DB	R2 Test **in** DB	R1 Test **not in** DB	R2 Test **not in** DB
Aedes aegypti -Aae IAEA (15M, 15F)	100% Aae IAEA	100% Aae Strains (53% UG, 47% LSHTM)	96% Aae Strains (72% UG, 24% LSHTM), 4% Aalbo	88% Aae Strains (62% LSHTM, 26% UG), 12% Aalbo
Aedes aegypti – Aae LSHTM (15M, 15F)	100% Aae Strains (80% LSHTM, 16% UG, 4% IAEA)	100% Aae Strains (84% UG, 16% IAEA)	100% Aae Strains (73% UG, 27% IAEA)	80% Aae Strains (54% IAEA, 26% UG), 7% Mu, 13% Coq p
Aedes aegypti – Aae UG (15F)	100% Aae Strains, (80% UG, 20% LSHTM)	100% Aae Strains (55% LSHTM, 20% Aae UG, 25% Aae IAEA)	100% Aae (45% IAEA, 55% LSHTM)	75% Aae Strains (38% LSHTM, 37% IAEA), 12% Cq Mu, 13% Coq p
Aedes vexans -Avex Atl (30F).	97% Avex Strains (51% Atl 40% UG, 6% Den), 3% Aalbo	85% Avex Strains (45% UG%, 40% Den), 15% Aae LSHTM	85% A vex Strains (45% Den, 40% UG) 15% Aae LSHTM	75% Avex Strains (35% Den, 40% UG), 25% Aae LSHTM
Aedes vexans – Avex Denver (30F)	97% Avex Strains (44% Atl, 33% Den, 20% UG), 3% Aae LSHTM	82% Avex Strains (34% UG, 20% Den, 28% Atl), 9% Aae LSHTM, 9% Otaen	96% Avex Strains (53% Atl, 43% UG), 4% Aae LSHTM	67% Avex Strains (38% Atl, 29% UG), 14% Otaen, 19% Aae
Aedes vexans – Avex UG	100% Avex Strains (31%	100% Avex Strains (62%	83% Avex Strains (55% Atl, 28%	61% Avex Strains (33%Atl, 28% Den),

(15F)	Den, 54% UG 15% Atl).	Atl, 23% Den, 15% UG)	Den) 14% Aae, 3% Otaen	16% Aae, 23% Otaen
Aedes albopictus - Aalbo IAEA (15M, 15F)	92% Aalbo Strains (77% IAEA, 15% Sav) 8% AaeUG	75% Aalbo Strains (67% Sav, 8% IAEA), 17% Aae, 8% Otriseri	75% Aalbo Strains (54% Sav, 21% UG), 10% Aae 15% Otriseri	39% Aalbo strains (31% Sav, 8% UG), 38% Aae, 23% Otriseri
Aedes albopictus University of Georgia – Aalbo UG (16F)	100% Aalbo Strains (100% UG)	81% Aalbo Strains (81% Sav) 19% Aae	81% Aalbo strains (62% IAEA, 19% Sav), 19% Aae	56% Aalbo Strains (36% IAEA, 20% Sav), 18% Avex, 26% Aae
Aedes Albopictus Savannah - Aalbo Sav' (10F)	100% Aalbo Strains (80% Sav 20% IAEA)	80% Aalbo Strains (60% UG, 20% Sav), 20% Aae	80% Aalbo Strains (50% IAEA, 30% UG), Avex 20%	40% Aalbo strains (25% AEA, 15% UG), 38% Aae, 22% Avex
Anopheles arabiensis - Aarab, IAEA (8M, 8F)	100% Aarab Strains (100% IAEA)	77% Aarab Strains (77% ICL), 15% Agam G3, 8% Agam Tia (**100% Agam Complex**)	76% Aarab Strains (76% ICL) 12% Agam Tia, 12% Agam G3 (**100% Agam Complex**)	40% Agam G3, 20% Agam Tia, 20% Keele, 20% Keele **(100% Agam Complex)**
Anopheles arabiensis – Aarab ICL (7M, 7F)	100% Aarab Strains (75% ICL, 25% IAEA)	75% Aarab IAEA, 17% Agam G3, 8% Ngu **(100% Agam Complex)**	75% Aarab (IAEA) 17% Agam G3, 8% Ngu **(100% Agam Complex**)	84% Agam Tia, 8% Agam G3, 8% As STML **(92% Agam Complex)**

Anopheles gambiae – Agam G3 LSHTM (15M, 15F)	100% Agam G3 LSHTM Strains	39% Agam Kisumu, 29% Keele, 4% Agam Tia, 26% Agam Ngu, 2% As Red **(98% Agam complex Strains)**	14% Agam Kisumu 36% Keele, 36% Aarab, 7% Tia, 7% As Dub5 **(93% Agam Complex)**	27% Keele, 19% Agam Kisumu,12% Agam Ngu, 12% Agam Tia, 8% Aarab, 3% As Dub5, 10% As Red, 9% As Beech **(78% Agam Complex)**
Anopheles gambiae – Agam G3 ICL (10F, 10M)	71% Agam G3 ICL Strain, 29% Keele **(100% Agam complex)**	7% Agam G3 ICL, 39% Keele, 39% Aarab, 8% Agam Tia, 7% As Dub5 **(93% Agam Complex)**	72% Keele, 14% Aarab, 7% Agam Tia, 7% As Beech **(93% Agam complex Strains)**	72% Aarab, 21% As Dub5, 7% As STML **(72% Agam Complex Strains)**
Transgenic Cas9 – Trans (gene edited Agam G3 ICL), (10M, 10F)	100% Trans	67% G3 ICL, 33% Aarab **(100% Agam Complex Strains)**	67% G3 ICL, 33% Aarab **(100% Agam Complex Strains)**	13% G3 ICL, 27%, Keele, 42% Aarab, 8% Dub5, 10% As STML **(82% Agam Complex Strains)**
Anopheles gambiae – Agam Kisumu (20M, 20F)	100% Agam Kisumu	25% Aarab, 35% Keele, 8% Ngu, 10% Tia, 20% G3 (LSHTM), 2% Dub5 **(88% Agam Complex Strains)**	30% Aarab, 28% G3 (LSHTM), 30% Keele, 3% Ngu, 6% Tia, 3% Dub5 **(97% Agam Complex strains)**	36% Aarab, 36% G3 (LSHTM), 9% Keele, 7% Tia, 12% Ngu **(100% Agam Complex strains)**
Anopheles gambiae – Agam Tia	100% Agam Tia	21% Aarab, 26%, Keele, 16% Kis, 33%	27% G3, 29%, Aarab, 18% Kis, 26% Ngu **(100%**	22% G3, 25% Aarab 20% Keele, 33% Kis **(100% Agam**

(Tiassale) (15M, 15F)		Ngu, 4% G3 **(100% Agam Complex)**	**Agam Complex)**	**Complex)**
Anopheles gambiae N'guesso LSHTM (A coluzzii hybrid) (15M, 15F)	100% Agam Ngu LSHTM	76% Ngu ICL, 14% Keele, 3% G3 7% Dub5 **(93% Agam Complex)**	60% Ngu ICL, 33% Keele, 3% As Red, 4% As Beech **(93% Agam Complex)**	20% Ngu ICL, 14% Aarab, 21% Tia, 20% Keele, 18% As Red, 7% As Beech **(75% Agam Complex)**
Anopheles gambiae N'guesso ICL (A coluzzii hybrid) (10M, 10F)	100% Agam Ngu ICL	53% Ngu LSHTM, 39% Keele, 4% G3, 4% Tia **(100% Agam Complex)**	65% Ngu LSHTM, 35% Keele **(100% Agam Complex)**	16% Ngu LSHTM, 22% Tia, 13% As Red, 13%, Aarab Imp, 13% Keele 13% As STMl, 10%, As Red **(77% Agam Complex)**
Anopheles gambiae Keele (A coluzzii hybrid) (15M, 15F)	82% Keele, 9% Ngu, 9% Aarab **(100% Agam Complex)**	18% Keele, 45% Ngu, 17% G3, 5% Dub5, 15% Aarab **(95% Agam Complex)**	50% Ngu, 23% Aarab, 27% G3 **(100% Agam Complex)**	15% Ngu, 40% Aarab, 14% G3, 18% Tia, 8% Dub5, 5% As Red **(87% Agam Complex)**
Anopheles stephensi - As SDA500 LSHTM (16M, 16F)	82% As SDA500 6% As Red, 6% As Beech, 6% As STML, **100% As strains**	40% As Beech, 12% As Dub5, 36% As Red, 12% As STML, **100% As Strains**	38% As Red, 26% As Beech, 36% As Dub5, **100% As Strains**	40% As Beech, 12% As Dub5, 36% As Red, 12% As STML, **100% As Strains**
Anopheles stephensi - As Red LSHTM	100% As Red	34% As SDA500, 44% As Beech, 22%	35% As SDA500, 42% As Beech, 15% As Dub5,	44% As Beech, 34% As SDA500, 22% Agam Ngu, **78% As**

(15M, 15F)		As Dub5 **100% As Strains**	8%Ngu **92% As Strains**	**Strains**
Anopheles stephensi - As Beech LSHTM (15M, 15F)	60% As Beech 30% As Red, 10% As SDA500 **100% As strains**	31% As Beech, 22% As SDA500, 39% As Red, 8% Aarab. **92% As Strains**	23% As SDA500, 61% As Red, 8% As STML, 8% Aarab. **92% As Strains**	45% As SDA500, 31% As Red, 8% As Dub5, 8% Aarab, 8% Agam Ngu. **84% As Strains**
Anopheles stephensi – As STML LSHTM (15M, 15F)	87% As STML, 13% As Red. **100% As Strains**	3% As STML, 40% As Red, 35% As SDA500, 22% As Beech. **100% As Strains**	40% As SD500, 22% As Beech, 23% As Red, 3% Agam Kis, 4% G3, 8% Tia. **85% As Strains**	44% As SDA500, 25% As Beech, 15% Agam Kis, 16% Agam G3. **69% As Strains.**
Anopheles stephensi – As Dub5 (15M, 15F)	67% As Dub5, 33% As Red. **(100% As Strains)**	42% As Red, 35% As Beech, 16% As SDA500, 7% As STML **(100% As Strains)**	42% As Red, 35% As Beech, 16% As SDA500, 7% As STML **(100% As Strains)**	25% As Beech, 18% SDA500, 30% As Red, 18% Agam Kis, 9% Agam Ngu **(73% As Strains)**
Coquillettidia perturbans Coq p UG (16F)	100% Coq p. (32% UG, 56% Den, 12% Atl)	100% Coq p (31% Atl, 31% Den, 32% UG, 6%, Sav)	94% Coq p (56% Atl 38% Sav), 6% Avex	92% Coq p (44% Atl, 48% Sav), 8% Avex
Coquillettidia perturbans Coq p Sav (14F)	100% Coq p (Sav 59%, 41% UG)	100% Coq p (25% Sav, 25% Den, 25% Atl, 25% UG)	100% Coq p (42% Atl, 42% Den 16% UG)	100% Coq p (42% Atl, 42% Den, 16% UG)
Coquillettidia perturbans Coq p Den	100% Coq p (Den)	100% Coq p (72% Atl 14%, Sav 14%	100% Coq p (86% UG, 14% Sav)	100% Coq p (72% Atl 14% Sav, 14% UG)

(15F)		UG)		
Coquillettidia perturbans Coq p Atl (15F)	100% Coq p (72% Atl, 14% UG, 14% Den)	100% Coq p (6% Atl, 35% UG, 45% Den, 14% Sav)	100% Coq p (43% Den, 21% Sav, 36% UG)	100% Co q (29% Sav, 42% Den, 29% UG)
Culex quinquefasciatus Cq Muheza LSHTM (15M, 15F)	100% Cq Strains (71% Mu, 21% TPRI, 8% Sav)	71% Cq Strains (29% Mu, 42% TPRI), 29% Coq p Atl	71% Cq Strains (44% TPRI, 27% Sav), 15% Coq p, 14% Avex	33% Cq Strains (16% TPRI 17% Sav), 13% Cq tar, 13% Aae, 25% Avex 8% Coq p 8%
Culex quinquefasciatus TPRI (10M, 10F)	100% Cq Strains (80% TPRI 20% Mu)	100% Cq Strains (20% TPRI, 58% Mu, 22% Sav)	84% Cq Strains (38% Mu, 46% Sav), 11% Avex, 5% Aae	48% Cq Strains (20% Mu, 28% Sav) 18% Coq p, 20% Aae, 14% Avex
Culex quinquefasciatus Savannah (5F)	100% Cq Strains (Sav 67% Mu 22% TPRI 11%)	66% Cq Strains (44% TPRI, 22% UG) 22% Aae, 12% Avex	66% Cq Strains (22% UG, 22% Mu, 22% TPRI) 24% Aae, 10% Avex	55% Cq Strains (22% Ug, 22% Mu, 11% TPRI), 12% Avex, 33% Aae
Culex quinquefasciatus UG (5F)	100% Cq strains (50% Sav, 25% Mu, 25% UG)	75% Cq Strains (50% Sav, 25%, UG) 25% Avex	85% Cq Strains (50% TPRI, 10% Mu 25% Sav), 15% Avex	40% Cq Strains (20% TPRI, 20% UG), 30% Avex, 30% Aae
Ochlerotatus taeniorhynchus – Otaenor Savannah (6F)	100% Otaenor Strains (80% Sav, 20% UG)	70% Otaenior Strains (40% UG, 30% Sav) 20% Avex, 10% Aae	60% Otaenior Strains (60% UG), 30% Avex, 10% Aae	35% Otaenior strains (35% UG), 40% Avex, 25% Aae
Ochlerotatus taeniorhynchus UG (6F)	100% Otaenior (UG 100%)	55% Otaenior (55% Sav), 33% Otriseri 12%	55% Otaenior (Sav 55%) 33% Otriseri 12% Avex	25% Otaenor (25% Sav), 18% Otrisei, 57% Avex

		Avex		
Ochlerotatus triseriatus Otriseri Savannah (6F)	100% Otriseri Sav	60% Otriseri (60% UG) 25% Otaenior, 15% Aae	60% Otriseri (60% UG), 25% Otaenior 15%, Aae	50% Avex, 50% Aae, 0% Otriseri
Ochlerotatus triseriatus UG (6F)	100% Otriseri (UG 100%)	43% Otriseri Strains (43% Sav), Avex 57%	43% Otriseri Strains (43% Sav), 57% Avex	72% Avex, 14% Otaenor, 14% Aae, 0% Otriseri
Ochlerotatus sollicitans Osoll UG (7F)	100% Osoll UG	77% Osoll Sav, 23% Avex	77% Osoll Sav, 23% Avex	29% Osoll Sav, 29%, Avex, 13% Otriseri 29% Otaenior
Ochlerotatus sollicitans Sav (7F)	100% Osoll Sav	100% Osoll UG	100% Osoll UG	60% Aae lshtm, 40% Otaenor, 0% Osoll
Psorofera ferox – Pferox Savannah (6F)	100% Pferox Strains (80% UG, 20% Sav)	100% Pferox Strains (Sav 80%, UG 20%)	80% Pferox Strains (80% UG) 20% Aae lshtm	40% Otaenor, 60% Aae lshtm, 0% Pferox
Psorofora ferox UG (6F)	100% Pferox UG	100% Pferox Sav	100% Pferox Sav	40% Avex Den, 40% Otriseri, 20% Otaenior, 0% Pferox

M = Numbers of males used, F = Numbers of females used; DB= database.

R1 & R2 = Rank 1 and Rank 2 correctly identified strains

Notes:

1. Short forms of the species names and the locations of their collections were used across the rows and in subsequent tables below and are the same as in Table 1. In the case of *A. gambiae,* sibling species were included in the total of correct strains identified.

2. There was only one reference image of each strain in the database and once this had been selected at rank 1, it was not possible for it to be selected again at rank 2, hence the 0% correct strains at rank 2 when rank 1 had 100% correctly identified strain results (no further correct strains could be selected at rank 2).

3. In cases where strains were part of a species complex, the total percentage of sibling strains/complexes selected is highlighted in bold.

Table 3: The difference between the percentages of correctly identified strains in the presence and absence of the test strain in the database.

Strain	Rank 1 test in DB	Rank 1 test not in DB	Rank 1 P values	Rank 2 test in DB	Rank 2 test not in DB	Rank 2 P values
Aae IAEA (30)	100%	96%	0.24043	100%	88%	0.0523
Aae LSHTM (30)	100%	100%	No Difference	100%	80%	**0.0105**
Aae UG (15)	100%	100%	No Difference	100%	75%	**0.0418**
Avex Atl (30)	97%	85%	0.1073	85%	75%	0.3370
Avex Den (30)	97%	96%	0.834	82%	67%	0.1863
Avex UG (15)	100%	83%	0.1007	100%	61%	**0.0080**
Aalbo IAEA (30)	92%	75%	0.0786	72%	39%	**0.0108**
Aalbo UG (15)	100%	81%	0.0810	81%	56%	0.1473
Aalbo Sav (10)	100%	80%	0.1462	80%	40%	**0.0230**
Aarab IAEA (16)	100%	100%	No difference	100%	100%	No difference
Aarab ICL (14)	100%	100%	No difference	100%	92%	0.2888
Agam G3	100%	93%	0.1435	98%	78%	**0.0181**

LSHTM (30)						
Agam G3 ICL (20)	100%	93%	0.1435	93%	72%	0.0844
Transgenic ICL (20)	100%	100%	No difference	100%	82%	0.5998
Agam Kis (40)	100%	97%	0.2727	88%	100%	**0.0247**
Agam Tia (30)	100%	100%	0.0905	100%	100%	No difference
Ngu LSHTM (30)	100%	93%	0.1435	93%	75%	0.0593
Ngu ICL (20)	100%	100%	No difference	100%	77%	**0.0244**
Keele Ugl (30)	100%	100%	No difference	95%	87%	0.2830
As SDA LSHTM (32)	100%	100%	No difference	100%	100%	No difference
As Red LSHTM (30)	100%	92%	0.1169	100%	78%	**0.0069**
As Beech LSHTM (30)	100%	92%	No difference	100%	84%	0.0235
As STML LSHTM (30)	100%	100%	No difference	100%	69%	**0.0010**
As Dub5 (30)	100%	100%	No difference	100%	73%	**0.0024**
Coq p UG (16)	100%	94%	0.3275	100%	92%	0.2557
Coq p Sav (14)	100%	100%	No difference	100%	100%	No difference
Coq p Den (15)	100%	100%	No difference	100%	100%	No difference
Coq p Atl (15)	100%	100%	No	100%	100%	No

			difference			difference
Cq Mu (30)	100%	81%	**0.0128**	71%	33%	**0.0149**
Cq TPRI (20)	100%	84%	0.0655	100%	48%	**0.0002**
Cq Sav (12)	100%	66%	**0.0300**	66%	55%	0.5895
Cq UG (8)	100%	85%	0.2701	75%	40%	0.1704
Otaenior Sav (7)	100%	60%	0.0714	70%	35%	0.2064
Otaenior UG (7)	100%	55%	0.0520	55%	25%	0.2696
Otriseri Sav (6)	100%	60%	0.0973	60%	0%	0.0299
Otriseri UG (6)	100%	43%	**0.0363**	43%	0%	0.0826
Osollici Sav (7)	100%	77%	0.1937	77%	29%	0.0830
Osollici UG (6)	100%	100%	No difference	100%	0%	Not applicable
Pferox Sav (6)	100%	80%	0.2689	100%	0%	Not applicable
Pferox UG (6)	100%	100%	No difference	100%	0%	Not applicable

The full names of the strains and locations collected are the same as in Tables 1, 2 & 4. The numbers in brackets next to the strains indicate the total number of strains tested. The comparison of proportions was calculated using Med Calc at the 95% confidence interval and 0.05 significance level - Med Calc https://www.medcalc.org/calc/comparison_of_proportions.php. Significant values ($p < 0.05$) are in bold in the above table.

Not applicable = As there were only 2 similar strains in the database, each strain with just one wing image in the database, when one of the strains was removed it would leave just one wing image, which when selected at rank 1 would leave no further

images of the correct strain to be ranked 2 from the database. Hence the 0% result at rank 2 when 100% of the one remaining strain had been ranked at 1.

Table 4. Observations for each species and strain with the test strain either in or out of the database.

Species and Strains	Observations
Aedes aegypti IAEA, LSHTM, UG	All three strains of *A. aegypti* selected for each other with and without the test being present in the database. Anophelines were never selected in the early ranks (ranks 1 and 2, with or without the test strain in the database), *Culex* and other *Aedes* species were, especially when the correct strain had already been ranked 1 or 2 or the test was absent from the database. Field caught *A. aegypti* UG (from the US) selected for laboratory reared *A. aegypti* LSHTM (from West Africa), suggesting that phenotypic characteristics remained constant even when this species was laboratory reared over many generations. Both laboratory reared strains selected more for the field caught strain UG (53%, 84% UG, rank 2 test in the database), rather than the laboratory strains indicating a greater resemblance to the field strains. *Aedes aegypti* are known to have arrived in the US from West Africa (Powell and Tabachnick, 2013) and this selection by the field caught US strain for the West African strain reared in the laboratory by the LSHTM, suggests that the phenotypic characteristics of the US strain has remained constant with the strain from which it originated. The IAEA strain which originated from Brazil and was reared in the laboratory in Austria also selected for the other 2 strains, this too would have arrived to the New World (the continent of North and South America) in ships from West Africa (Powell &Tabachnick, 2013). Living far away from its country of origin, over the centuries, has not changed the physical characteristics of *Aedes aegypti* originating from West Africa.

Aedes vexans Atl, Den, UG	*Aedes vexans,* the inland floodwater mosquito is cosmopolitan and a common pest species of medical importance. All three strains, field collected from different places in the US, largely Atlanta and Denver, selected for each other when the test was not present in the database. The Denver strain, collected at 4,995 feet and the Atlanta strains (general elevation 1,050 feet), both selected for each other when the exact strain was absent from the database, 93% (Denver absent), 85% (Atlanta absent) and 83% (UG, Atlanta absent) of the time at rank 1. This indicates that the geometry of the wing veins and the shape of the wing in all three strains is similar, suggesting that there is no phenotypic divergence of the species collected at the higher elevations from Denver. *A. vexans* is less prevalent at higher elevations and absent from elevations above 8,000 feet (Harmston & Lawston, 1967).
Aedes albopictus IAEA, UG, Sav	*Aedes albopictus,* native to Asia and known as the 'Asian tiger mosquito' has rapidly invaded many continents including Africa, Europe and both North and South America, largely via the trade in used tyres from Asia where they often breed (Ngoagouni et al 2015). The results show that the laboratory reared strain (IAEA) originating from Rimini Italy, selected for both field caught strains from the US and vice versa. A greater percentage of field caught strains were selected by the laboratory reared strains (64% rank 2 test in the database; 75% rank 1 test not in the database), indicating that laboratory strains selected and therefore resembled field strains more than they did other laboratory reared strains in the early ranks, 1 and 2. The physical characteristics of the wings appears not to have changed a great deal as this species moved across continents. 77% of the laboratory reared strains selected for itself at rank 1; 15% were field caught *A. albopictus* Sav (when a copy of the lab strain was present in the database). Whilst this is high, it suggested that there may be some difference between the lab and field caught strains, but not

	enough that they could not select for the field caught strain when a copy of the lab reared strain was absent from the database when 75% of the field caught strains were selected and ranked 1. The field caught strains selected for each other rather than the lab strain at rank 1 (when present in the database), selecting the lab strain at rank 2 and when a copy of the test was not in the database. As with *A. aegypti* (above), the *A. albopictus* tested here, have largely remained unchanged as the species crossed continents, at least in terms of wing morphology. Once the correct strains had been ranked one and two, subsequent rankings were of other *Aedes* species. As there was only one copy of each strain in the database, which when selected at rank 1 or 2, could not be selected again, only images closely resembling the test (i.e. other *Aedes* species) could be selected after a rank 1 or 2 correct strain selection.
Anopheles arabiensis IAEA, ICL	Both laboratory reared strains (one obtained from MR4 Dongola, reared in the UK at ICL; the other reared in Austria (IAEA) and also originating from Dongola, Sudan, Africa, selected for each other when the test was absent from the database. In rank 2, members from the *A. gambiae* complex were selected indicating the close resemblance of *A. arabiensis* to its siblings within the *A. gambiae* family. The difference between the members of the complex was sufficient for I^3S to distinguish between them when a copy of the test was in the database. As before (Vyas-Patel & Mumford, 2017), I^3S distinguished between *A. arabiensis* from other *A. gambiae* complex members, 100% of the time at rank 1 when a copy of the test was in the database. When it was not present in the database, closely resembling siblings from the *A. gambiae* complex family were selected at ranks 1 and 2, illustrating the importance of knowing the contents of the database when dealing with sibling species. The strain from Dongola is used for the Sterile Insect Technique (SIT) at IAEA and at the time of carrying out the study, the origin of the ICL

	strain was not known (it was later found to also originate from Dongola, after the test had been completed), indicating that I^3S accurately selects for similar strains and species when present in the database, without human bias.
Anopheles gambiae G3 LSHTM, ICL	*Anopheles gambiae* G3, colonised in London in 1975, originates from McCarthy Island (The Gambia, West Africa). The G3 reared at ICL was obtained from BEI Resources, MR4 (The Malaria Research and Reference Reagent Resource Centre, https://www.beiresources.org/MR4Home.aspx), the origin of the G3 from LSHTM was not known. There was 100% recall of G3 from LSHTM at rank 1, but not so for the G3 ICL strain where it was 71% at rank 1 with a test sample in the database, the remaining selection was largely for the Keele strain (known to be the 'M' form, *A. coluzzii*) and some *A. arabiensis* (ICL). Furthermore, the G3 from LSHTM did not select for ICL G3 and vice versa at all, at any stage (rank 1 & 2 with or without a test sample in the database), suggesting that they did not resemble each other enough for rank 1 & 2 retrieval. Both G3s (from ICL & LSHTM) selected for and therefore resembled the Keele (originally a hybrid of *A. gambiae* and *A. coluzzii,* now designated as the M form, *A. coluzzii;* Ranford-Cartwright et al, 2016) and Nguesso strains (which has also been reclassified as the M form, *A. coluzzii;* Habtewold et al, 2016). These were selected at rank 1 (in the case of ICL G3) and rank 2 (in the case of LSHTM G3). Keele and N'guesso strains were also selected at rank 2 for both G3 strains indicating the close similarity between members of the

	A. gambiae complex and reflecting their hybrid origins. The molecular markers (Lee et al 2014) from the analysis of 24 individuals of G3, of the Imperial College strain (ICL), was found to be a mix of *A. coluzzii* (M form) and *A. gambiae* s.s. (S form). The ICL strain was a hybrid, the X chromosome retained the mix of the IGS markers used to distinguish both the M & S forms (Favia et al 2001) and the colony was polymorphic for the 2La inversion. In contrast, random and repeated testing of G3s from the LSHTM colony indicated that the individuals tested were *A. gambiae* s.s. (personal communication from lead researchers at both ICL (Dr Nikolai Windbichler) and LSHTM laboratories (Miss Shahida Begum and Prof James Logan). Thus the two colonies of G3 (ICL and LSHTM) were genetically different. This was not known at the time of testing and I^3S was able to distinguish between them. Despite the expectation that both G3 strains would select for each other they did not, as genetically, they were different and this difference was reflected in the wing morphology and picked up by the software (I^3S). Vector Base, a repository for different insect species and strains, stresses the hybrid nature of G3 and the need to take care when using G3 colonies (https://www.vectorbase.org/).
Anopheles gambiae G3, Transgenic Cas9.	The Cas9 transgenic mosquito strain, targeting female reproduction was created from *Anopheles gambiae* G3, ICL (Hammond et al, 2016) and repeatedly selected for fluorescent marker expression in subsequent generations. It retrieved itself 100% at rank 1 when it was present in the database. In rank 2, the selection was 67% for G3 ICL from which it was derived (never G3 LSHTM) and the rest was for *A. arabiensis* (33%). As there was only one copy of each sex of this strain, it could only select for itself **once** for each test specimen and in this case it was ranked 1, 100% of the time, indicating that I^3S could be a useful tool for detecting transgenic strains and probably other highly selected strains at least in the early stages after release.

	Phenotypic traits could change over several generations in the field due to hybridisation after release and this would need to be studied. When the strain was absent from the database, it selected for other members of the *A. gambiae* complex to which it was similar, but it should always be tested for with a copy of the original transgenic strain present in the database; without it, the test would not give meaningful results.
Anopheles gambiae Kisumu (Kis)	*A. gambiae* from Kisumu, East Africa is a strain which has been laboratory reared in London since 1998 and is susceptible to pyrethroids. It was fully retrieved at rank 1 when an image of the test was in the database (100%). At rank 2 and when the test was absent, other members of the *A. gambiae* members were selected, especially *A. arabiensis* and G3 from LSHTM. *A. gambiae* is sympatric (occurs together) with *A. arabiensis* especially in areas such as Kisumu, East Africa; both are important vectors of malaria in Africa. Of the seven closely related sibling species comprising the *A. gambiae* complex, these two species are the most widespread, occurring in sympatry throughout Sub Saharan Africa and nearby islands. Despite this, the incidence of naturally occurring interspecies hybrids (recognized chromosomally) is at or below 0.2% (Coluzzi et al, 1979). Clearly the ability to distinguish between the two sibling species using I^3S as demonstrated here, gives rise to a very important and easy to use, non-molecular tool. However, when testing, images of both sibling species need to be in the database. Otherwise only the next closely resembling species/strain from the database could be ranked 1.
Anopheles gambiae Tiassale (Tia)	Colonised at LSHTM in 2013, *A. gambiae* Tiassale originates from Tiassale, a town in southern Ivory Coast, West Africa. After selecting itself 100% at rank 1 (test in the database), it selected for other members in the *A. gambiae* complex 100% of the time in each case. There was only one strain of Tiassale from one location and there was only one wing copy of each sex of

	Tiassale in the database, which when selected at rank one, could not be selected again. Hence the 0% *A. gambiae* Tiassale after the rank 1 selection.
Anopheles gambiae N'guesso LSHTM & ICL (*A. coluzzii*)	*Anopheles gambiae*, strain N'guesso was colonised in 2006 and originates from Yaounde, Cameroons. It was designated as the 'M' form of *A. gambiae* (Habtewold et al 2016). The 'M' form has now been renamed as a separate sibling species within the *A. gambiae* complex and is known as *Anopheles coluzzii* (Coetzee et al 2013). Both LSHTM and ICL strains were correctly selected at rank 1, 100% of the time when the test strain was present in the database and also selected for each other at rank 2, test sample present in the database. They continued to select for each other at ranks 1 & 2 when the test was absent from the database, indicating the close similarity between the strains reared at different places. The LSHTM strain was obtained from ICL in 2013. However, this was not known at the time of carrying out the test, once again indicating the complete lack of human bias in the testing and the accurate results from I^3S.
Anopheles gambiae Keele (*A. coluzzii*)	The Keele colony was from the University of Glasgow. Molecular characterisation indicated that it was the 'M' form of *A. gambiae* (*A. coluzzii*) despite originating from a mixed colony of 'M' (*A. coluzzii*) and 'S' (*A. gambiae* s.s.) strains (Ranford-Cartwright, et al 2016). It selected itself 82% of the time at rank 1 and 18% at rank 2 (test present in the database), the remaining selection was for *A. gambiae* N'guesso (9%) which is also the 'M' form/*A. coluzzii* and 9% *A. arabiensis*, a sibling species and part of the *A. gambiae* complex family. At rank 2, the N'guesso strain (*A. coluzzii*) strain was once again selected (45%) and it continued to be selected when the test was absent from the database, reinforcing the similarity between Keele and N'guesso which have both been designated as the 'M' form *A. coluzzii,* originating from a mixed colony. Later selections were for other members from the *A. gambiae* family. *Aedes* and *Culex* were never selected.

Anopheles stephensi SDA500, LSHTM	Originating from Sind province, Pakistan, *A. stephensi* SDA500 was reared at the LSHTM for studies on the susceptibility of the strain to malaria parasites. Rank 1, with the reference image in the database, resulted in the majority of selections for SDA500 (82%), followed by *A. stephensi* Red, Beech and STMAL (all 6%). Subsequent selections at rank 2 and when the test was absent from the database were for the other *A. stephensi* strains and as there were five different strains of *A. stephensi* present in the database, it was possible to select for them in greater numbers than if there were fewer strains present. The result was 100% selection (in total) for the other *A. stephensi* strains.
Anopheles stephensi Red, LSHTM	This strain had genetically recessive Red eyes and was a laboratory colony. It selected for itself 100% at rank 1 with the reference image in the database, unlike the other *A. stephensi* strains which selected for each other at rank 1 as well as for itself when tested. This indicates that it is sufficiently distinct from the other *A. stephensi* strains in the database, but not so different that it could not select for the correct strain at rank 2 and when the test was absent from the database when large numbers of the other *A. stephensi* strains were selected (SDA500, Beech and Dub5).
Anopheles stephensi Beech, LSHTM	Originating from New Delhi, India, laboratory reared since 1947 and from a colony held at Beecham's laboratory from the 1950s, this strain was susceptible to insecticides and indicated high infectivity to the malaria parasite. It selected the other strains of *A. stephensi* as well as itself at rank 1, test in the database. It continued to select for the other strains of *A. stephensi* when the test was absent from the database at high levels (84% to 92%), indicating the morphological similarity between all the *A. stephensi* strains present in the database.
Anopheles stephensi STMAL, LSHTM	The strain *A. stephensi* STMAL, so named due to its resistance (both in the larval and adult stages) to the organophosphate insecticide Malathion, originates from Lahore, Pakistan. It

	selected for itself 87% and the *A. stephensi* Red strain 23% at rank 1, test in the database. Selections when the test was absent from the database at rank 1 were for SDA500 (also from Pakistan), Beech (from India) and *A. stephensi* Red.
Anopheles stephensi Dub 5	Originally from Dubai this strain was reared for studies on the susceptibility to insecticides as it was resistant to pyrethroids. Colonised at the LSHTM since 1989. As with STMAL, it selected itself (67%) and *A. stephensi* Red (33%) at rank 1, test in the database. Later selections, rank 2 test in the database and rank 1 test not in the database, resulted in 100% selection in total, of the different *A. stephensi* strains present in the database.
Coquillettidia perturbans UG, Sav, Den, Atl	*C. perturbans* is found across five continents and its range continues to grow. It is also the vector for a number of diseases most notably West Nile virus and Eastern Equine Encephalomyelitis. This field caught strain, from all three sites (Savannah, Denver and Atlanta) selected for itself and the other two strains at rank 1 and 2, test present in the database. When the test was absent from the database, the other two strains were selected at 92% to 100% levels, indicating the close similarities in appearance of the wings between all of the field caught strains.
Culex quinquefasciatus Muheza (MU), LSHTM	*Culex quinquefasciatus* found in tropical and sub-tropical areas around the world is an important vector of a number of major arboviruses and nematode parasites infecting man, other mammals and birds. From Muheza, Tanzania (East Africa), it was generally considered to be pyrethroid resistant. It selected for itself 71% and the other strain from Tanzania, TPRI at 21%, but also selected the field caught strain SAV 8% at rank 1 (test present in the database). It selected other *Culex* and *Aedes* strains later on in the ranks.
Culex quinquefasciatus TPRI, LSHTM	The insecticide susceptible strain *C. quinquefasciatus* from the Tropical Pesticide Research Institute (TPRI), Tanzania had been reared at the LSHTM since 2010. It selected itself 80% of the time and the Muheza (MU) strain 20% at rank 1, test strain present in

	the database. It also selected for the field caught *C. quinquefasciatus* from the US (SAV), indicating that the field caught strains from the US were very similar to the laboratory reared strains from Africa. Other mosquitoes that were selected later on were *Culex* and *Aedes* species, never Anophelines, indicating that I^3S segregates the genus of mosquitoes very well from wing images.
Culex quinquefasciatus UG, Sav	Found in the tropical and sub-tropical regions of the world, including most of southern United States and South America, this field collected strain from the US selected for all three strains present in the database at rank 1, test in the database (TPRI, Mu & field collected UG), indicating the uniformity and similarity of the morphology of laboratory reared and field caught strains. When it was not possible to select for the correct strains (test absent and other correct strains already selected), the test image selected for other *Aedes* and *Culex* species, never *Anopheles*.
Ochlerotatus taeniorhynchus (*Aedes taeniorhynchus*) Sav, UG	*Ochlerotatus* was previously known as *Aedes* until a name change occurred in 2000 and it may still be called *Aedes* by many workers and from prior records before the name change. *O. taeniorhynchus* is prevalent along the eastern coast of both North and South America, it is also known as the black saltmarsh mosquito and is a ferocious biter. The strains from both locations recalled themselves 100% at rank 1 and then retrieved each other from rank 2 (test present in the database). Other *Oclerotatus* and *A. vexans* were selected in later ranks, test absent from the database.
Ochlerotatus triseriatus (*Aedes triseriatus)* Sav, UG	This tree hole breeding mosquito is found along the eastern coast of North America. Commonly known as the eastern tree hole mosquito, it can breed in other containers such as hollows in discarded tyres and has the potential to spread via the trade in tyres across continents. Primarily zoophilic, it also bites humans, is the main vector for La Crosse virus and transmits a number of other viruses in the laboratory. The correct strain was retrieved in

	both cases, 100% at rank 1 (test in the database) and then retrieved each other at rank 2 and when the test was absent in the database. *Aedes vexans* and *Ochlerotatus taeniorhynchus* were retrieved when the test was absent in the database in later ranks (rank 2, test absent).
Ochlerotatus sollicitans (*Aedes sollicitans*) Sav, UG	*O. sollicitans*, known as the eastern saltmarsh mosquito is largely found along the eastern coast of the US and Canada and a few of the islands near the east coast (generally never the Pacific coast along the West). It can also be found in saline pools as far inland as Arizona. As well as being a nuisance species, it is an important transmitter of arboviruses in mammals (equine encephalitis, dog heartworm). Both strains (Sav & UG) selected for themselves 100% of the time at rank 1 (test in database). They selected for each other at rank 2 (test in database) and when one of the strains was absent from the database. When it was not possible to select for the correct species (test not in database or the correct strain had already been ranked 1), they selected for other *Ochlerotatus* and *Aedes* species (mostly *A. vexans*).
Psorophora ferox Sav, UG	*P. ferox*, known as the white footed woodlands mosquito and a ferocious biter, is capable of harbouring and transmitting a number of different arboviruses and is native to most woodland areas of North and South America. The results indicate that the strains from both areas select for each other even at rank 1, test present in the database (Sav strain selects 80% for the UG strain), indicating the similarity of the wings from both sites. When the test had been correctly selected at rank 1, (sample in the database), it selected the strain from the second location at 100% levels at rank 2. The selection for Sav/UG strains when one of the strains was removed from the database was also very high. When both strains had been logged at ranks 1 & 2, i.e. when it was not possible to log the correct strains (only one test sample of both sexes is in the database) and no further accurate samples of the test could be ranked (for example rank 2, test strain not in the

	database); *Aedes* species and *Ochlerotatus* species were selected. Although the sample numbers available were low, the results were 100% accurate.

IAEA = International Atomic Energy Agency, Vienna; **LSHTM** =London School of Hygiene & Tropical Medicine, UK; **UG** = University of Georgia, Atlanta, US. **ICL** = Imperial College London.

Discussion

I^3S can differentiate between different species of insects using wing images (Vyas-Patel & Mumford, 2017). The present investigation indicated that I^3S could also retrieve the correct species and strain if it was present in the database with 92% to 100% accuracy at rank 1 and select a different strain of the same species when the test strain was absent from the database (Tables 2, 3, 4). If the exact strain was absent from the database, closely resembling sibling species of the test strain would be selected provided they were present in the database (Table 2, *A. gambiae* species complex). I^3S is designed so that the image that most closely resembles the test is ranked first (i.e. rank 1), followed by the next most closely resembling image at rank 2 onwards. Concerns that differences in the morphology of strains from different places might affect accurate identifications were therefore unfounded.

Of the forty different strains tested - twenty field caught and twenty laboratory reared, three were significantly different at rank 1 when tested with and without the test specimen in the database (Table 3). In the laboratory reared *C. quinquefasciatus* Muheza (Cq Mu), the reason for the significant result, with a difference of only 19% (Table 3) was not clear. It could simply be different unknown factors that may affect laboratory reared species such as inbreeding, outbreeding (if it was originally a hybrid); or the different conditions, rigours and stress of continuous laboratory rearing in different places. It could be that the Muheza strain was different to the other *C. quinquefasciatus* strains and molecular tests could possibly shed further light on the differences seen here. However, molecular differences were unlikely as the significance level was low and it may simply be the effect of broken wings in one strain affecting the results in other strains. In the field caught *C. quinquefasciatus* Savannah strain (Cq Sav), the numbers were low, with broken wings, this could

contribute to the significant difference seen in Table 3, not just for Cq Sav but also for Cq Mu (and the other strains), which could not and did not select for Cq Sav in greater numbers. Similarly in the field caught *O. triseriatus*, where there were also low numbers and imperfect wings, giving rise to the significant result. However low numbers did not automatically mean that accurate results were affected. In the case of field caught *P. ferox* Sav and UG, both low in numbers (6 wings each) but with perfect, unbroken wing samples, the result was 100% selection of each strain for the other when one strain was taken out of the database (Tables 2, 3 and 4).

Of the rank 2 significant results (13 significant out of the 40 strains tested), the higher numbers of significant results were largely due to the fact that the pool of similar strains was twice reduced, firstly because the test strain had been removed from the database and secondly because many of the correct species and strains would already have been selected at rank 1 and therefore could not be selected again as there was only one copy of each strain in the database. Of the 13 significant results at rank 2, 10 were from the laboratory reared samples and only 3 were field caught strains. In this study, field caught strains were similar to each other compared to laboratory reared strains of the same species. Part of the reason for this could be the historical attempts by different workers to mix populations of a species collected from different regions into one cage, in order to prevent inbreeding during laboratory rearing, for example in the case of the Keele strain (Ranford-Cartwright, 2016).

A notable example of the above practice was seen in the results with the laboratory strain of *A. gambiae* G3, reared at both ICL and the LSHTM. The study was carried out with the expectation that the G3 strains from both places would select for each other, but not once did they do so, they were distinct from each other at both ranks 1 and 2. At the time of testing the genotype of the strains tested was not known. Subsequent enquiries revealed that previous molecular characterisation had indicated that the G3 from ICL was a hybrid of the M & S forms of *A. gambiae*, i.e. *A. coluzzii* and *A. gambiae* s.s. (Favia et al 2001) and that from the LSHTM was largely *A. gambiae* s.s. (Table 4). This indicated that the results from I^3S were fully in keeping with the results of molecular tests. In this study morphological variation followed genetic variation and was picked up by the software, I^3S.

A. stephensi, the other important vector for both *Plasmodium vivax* and *Plasmodium falciparum* malaria parasites, has a wide geographic range from the Middle East to India and continuing into China. Nonetheless, all of the strains tested here selected for each other, with and without the test in the database, with no significant differences at rank 1. Anophelines never selected for Culicines and vice versa indicating the accuracy of I^3S in selecting the correct genus in the early ranks (ranks 1 & 2).

There was very little strain difference between the samples collected from the field in the US. Any significant difference at rank 1, as occurred with *C. quinquefasciatus* (Sav) and *O. triseriatus* (UG), were probably due to broken wing samples affecting the analysis. The mosquito species collected from the field and tested here do not vary greatly (in terms of wing morphology) from place to place and a species from one area could easily select for the correct species from another area if the test was not in the database. Most notable for its high uniformity in appearance was *Coquillettidia perturbans* which was selected correctly across the board in all cases at 100% or very near 100% levels (92%).

Based on the results seen in this study, it should be possible to use I^3S to accurately differentiate between hybrid forms of *A. gambiae* in a collection of samples, also the 'M' & 'S' forms (*A. coluzzii* and *A. gambiae* s.s.) and to separate members of the *A. gambiae* complex with great accuracy. In order to do this, all forms of the *A. gambiae* complex, or at least those forms that were of interest, would need to be in the database, when the correct strain/sibling species was selected 100% of the time. It is no longer the case that '' the M & S forms of *A. gambiae* are morphologically indistinguishable and can only be identified using molecular techniques'' Lee et al (2014). I^3S and many other geometric techniques are capable of separating these sibling species and strains and should be used (Cañas-Hoyos et al 2014, Cao et al 2014 and 2015).

I^3S could also be used to differentiate transgenic or other highly selected mosquitoes from wild ones in the first phase after release, based on the results seen here, where transgenic *A. gambiae* were differentiated and retrieved 100% at rank 1 from the parent species from which they were genetically modified (here it was *A. gambiae* G3, ICL strain). This would require the prior testing of all wings, the transgenic or

otherwise selected form, the parent and the wild population wing images and the incorporation of all the relevant wing images into the database. The *A. gambiae* G3 parent tested here is a hybrid strain of *A. gambiae* and the results indicated that it could be distinguished from *A. gambiae* s.s (the LSHTM G3), which occurs in the field. Only one transgenic line was assessed here, so it should .not be assumed that every type of genetically modified mosquito could be similarly discriminated from the parent strain and each new transgenic line should be tested independently. As the parent strain was mostly selected at rank two (67% parent G3 at rank 2 when both transgenic and the parent G3 were in the database), there was a difference between them, albeit a small one. If the difference was large, the majority of the parent strains would have largely been selected beyond rank two, i.e. rank three, four or five. The reasons for any difference in wings between the transgenic and parent mosquitoes may be due to the effects from gene insertions or from selection in the insertion, rearing and maintenance process. Each generation is selected for with fluorescent marker expression, for instance. Furthermore, the exact mechanism and process could be different for any given transgenic mosquito produced, it might be subject to change over a few generations and would need further study. The long term identification potential after transgenic mosquito release is a conjecture, as it is not known what, if any changes could occur in wing morphology after many generations, over the long term, following hybridisation after any transgenic mosquito release.

Several studies (using different methods) have made distinctions between the shape of the wing, the size of the wing and the geometric pattern of the veins on the wing when considering the effects of the environment on morphological change in insect wings (Pieters et al 2017; Liu 2016; Yi Bai et al 2016 and 2015; Demari-Silva 2014; Prudhomme 2016; Sanzana et al 2013; Vidal & Suesdek 2012; Carvajal et al 2016). Here, I^3S takes into account the shape and vein pattern of the wings, as indicated by the red dots (Figure 1). The software requires the operator to mark the image. This creates a cloud of relevant points. I^3S only considers the relative positions and distances between these points. In this study, these points represent a set of features such as wing vein branches and wing shape, which are characteristic and present across the range of mosquito species and strains being examined. Apart from the marking of these landmark features on the wing image by the operator, all

of the calculations were achieved by the software to arrive at the closest match from the database and the ranking of all the images from the database, according to how closely they matched the test image (information contained on the I^3S website and associated tutorials http://reijns.com/i3s/, freely downloadable with the software, Hartog & Reijns 2013).

In this study the results relate to mosquito species, but there is every reason to believe that I^3S could just as effectively be used for other species and strains of insects. Using a variety of different geometric morphometric techniques has resulted in the differentiation of strains in other insect species (Cañas-Hoyos et al 2014, Cao et al 2014 and 2015). A comprehensive review of different geometric methods using mosquitoes as the study insect, described in detail how different geometric morphometric methods could be used to differentiate between insect species and strains, Lorenz et al (2017). The review stressed how all of the geometric morphometric methods reported thus far were capable of sorting not just species but also different strains of mosquitoes and other insects.

Every new species and strain should be tested independently, especially if different methods, software or tools are used. Information about the test species (where it was collected and when) should always be used and recorded. Features other than the wings should also be considered when testing any unknown species against a database of known species.

The results of image recognition software are ‘appearance based’. The software is designed to scrutinise and assess minute details of the markings made of an image, retrieve the closest match for a test from a database and rank the database images according to how closely they resemble the test. The software cannot ascertain the genotype of an organism, however, molecular differences were always reflected in the wings (at least in this study) for example the hybrid and non-hybrid *A. gambiae* G3 strains and could be discriminated and differentiated by the software. I^3S can sort, differentiate and identify different species and strains of a species with 100% accuracy as long as copies of the test are present in the database. It is a valuable complement to molecular characterisation and is an additional, useful tool to accurately identify species and strains.

One of the strongest reason to continue research into methods that use image recognition and geometric morphometrics techniques, is from a report of the National University of Colombia (Universidad Nacional de Colombia). The group from Colombia, Cañas-Hoyos et al (2014), described how two different strains of *Spodoptera frugiperda*; one that fed on corn and one on rice; could not be differentiated by any type of molecular method, yet they were able to differentiate between the two strains using wing morphometrics and stressed the importance of using the technique in the field. In the cohort of different species/strains examined here, molecular differences (if they existed) between species/strains were always reflected in the wings and could be picked up by the software, note the example of the laboratory reared *A. gambiae* G3 reared at different laboratories one of which was a hybrid and the other *A. gambiae* s.s. Both were differentiated 100% using I^3S when reference copies were in the database. Based on the examples mentioned above and all of the results seen in this study, it can safely be stated that any molecular difference between different species and strains are expressed in the phenotype of the wings and can be detected using image recognition such as I^3S or any one of the different geometric morphometric methods available currently.

The case for using image recognition and geometric morphometric tools to identify insect species and strains should not be ignored. These methods should be used alongside molecular characterisation and traditional taxonomic identification using keys. Will and Rubinoff (2004), urged that decisions taken in defining species and strains be based on ‘a constellation of data’ rather than just the narrow findings from any particular discipline, a conclusion supported here. Image recognition software for identifying insect species can clearly complement molecular methods and can be used effectively in conjunction with them or separately. Here, the image recognition software, I^3S has proven to be an accurate tool in the identification of different species and strains of insects, using their wing images.

Conclusion

1. A comprehensive study of a large number of different mosquito species and strains indicated that I^3S was accurately able to detect the differences not just between species but also between different strains of mosquitoes and rank them accordingly.

2. I^3S could accurately be used to differentiate between members of the *A. gambiae* complex, such as *A. gambiae* s.s., *A. coluzzii* and *A. arabiensis* with 100% accuracy when these sibling species were present in the database.
3. It is no longer the case that molecular characterisation provides the only way to identify sibling species of the *A. gambiae* complex. I^3S is capable of separating/identifying members of the complex with 100% accuracy.
4. In some cases of strain identification where it was impossible to identify insect strains using molecular characterisation, I^3S along with other geometric morphometric methods, could be able to differentiate these strains/sibling species and this should always be explored.
5. The presence of hybrid strains could be detected from either laboratory reared or field collected samples, provided a copy of the hybrid in question was present in the database.
6. The transgenic mosquitoes tested in this study could be detected and discriminated from the parent strains, suggesting highly selected lines of mosquitoes can be distinguished from parent strains.
7. I^3S can accurately be used to identify species and strains of medical importance and is a useful tool to keep watch on potentially invasive species/strains and their spread, such as *Aedes albopictus* from field collections.
8. Field specimens of the same species of mosquitoes collected from geographically different areas from the US vary little from each other and I^3S is capable of accurately selecting other strains of the test species if the test strain is absent from the database.
9. Laboratory reared strains can vary from different laboratories (as indicated by I^3S for example *A. gambiae* G3). Both morphological and molecular tests to identify the exact strain would be prudent before their use in experimental studies.
10. I^3S makes for an excellent additional and accurate tool to be used with both molecular and traditional identification using keys and should be utilised.
11. In future keeping copies of insect wings and other insect body parts, in databases, for use in image recognition of species/strain identification could become the norm.

Acknowledgements

Without the generous donation of insects from everyone mentioned in Table 1 (mosquito donors) this work would not have been possible. The interest and encouragement, correspondence and good wishes from every mosquito donor, was vital for the completion of this project. The correspondence with the I^3S team, Jurgen Hartog and Renate Reijns, was similarly invaluable. Thank you all!

References

Ampuero D Palma AT, Veliz D and Pardo LM, 2009. Description, seasonal morphological variation, and molecular identification of *Paraxanthus barbiger megalopae* obtained from the natural environment. *Heligoland Marine Research*. 64:172. https://doi.org/10.1007/s10152-009-0172-9. © Springer-Verlag and AWI 2009.

Ampuero D Palma AT, Veliz D and Pardo LM, 2010. Description, seasonal morphological variation, and molecular identification of *Paraxanthus barbiger megalopae* obtained from the natural environment. *Heligoland Marine Research*. 64:117–123. https://link.springer.com/article/10.1007/s10152-009-0172-9

Benítez HA, Lemic D, Bažok R, Gallardo-Araya CM, Mikac KM 2013. Evolutionary directional asymmetry and shape variation in *Diabrotica virgifera virgifera* (Coleoptera: Chrysomelidae): an example using hind wings. *Biological Journal of the Linnean Society*, Volume 111, Issue 1, Pages 110–118, https://doi.org/10.1111/bij.12194. https://academic.oup.com/biolinnean/article/111/1/110/2415800/Evolutionary-directional-asymmetry-and-shape.

Cañas-Hoyos N, E. J. Márquez and C. I. Saldamando-Benjumea, 2014. Differentiation of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) Corn and Rice Strains from Central Colombia: A Wing Morphometric Approach. *Annals of the Entomological Society of America* 107(3):575-581. 2014, https://doi.org/10.1603/AN12154.

Cao Y, Zhang C, Chen Q, Li Y, Qi S, Tian L, Ren Y, 2015. Identification of species and geographical strains of *Sitophilus oryzae* and *Sitophilus zeamais* using the visible/near-infrared hyperspectral imaging technique.*Pest Manag Sci*. 71(8):1113-21. doi: 10.1002/ps.3893. Epub 2014 Sep 25. https://www.ncbi.nlm.nih.gov/pubmed/25170577

Cao, Yang & Zhang, Chaojie & Chen, Quansheng & Li, Yanyu & Qi, Shuai & Tian, Lin & Ren, YongLin. (2014). Identification of species and geographical strains of *Sitophilus oryzae* and *Sitophilus zeamais* using VIS/NIR hyperspectral imaging technique. *Pest Management Science* 71. 10.1002/ps.3893. https://www.researchgate.net/publication/265172125_Identification_of_species_and_geographical_strains_of_Sitophilus_oryzae_and_Sitophilus_zeamais_using_VISNIR_hyperspectral_imaging_technique

Carreira VP, Soto IM, Hasson E & Fanara JJ, 2006. Patterns of variation in wing morphology in the cactophilic *Drosophila buzzatii* and its sibling *D. koepferae*. doi:10.1111/j.1420-9101.2005.01078.x. 1 9 *European Soc for Evol Biol*. 1275-1282. http://digital.bl.fcen.uba.ar/Download/paper/paper_1010061X_v19_n4_p1275_Carreira.pdf

Carvajal TM, Hernandez LF, Ho HT, Cuenca MG, Orantia BM, Estrada CR, Viacrusis KM, Amalin DM, Watanabe K, 2016. Spatial analysis of wing geometry in dengue vector mosquito, *Aedes aegypti* (L.) (Diptera: Culicidae), populations in Metropolitan Manila, Philippines. *J Vector Borne Dis.* 53(2):127-35 https://www.ncbi.nlm.nih.gov/pubmed/27353582 .

Coetzee M, Hunt RH, Wilkerson R, Della Torre A, Coulibaly MB, Besansky NJ, 2013. *Anopheles coluzzii* and *Anopheles amharicus*, new members of the *Anopheles gambiae* complex. *Zootaxa*. 2013; 3619:246–74. https://www.ncbi.nlm.nih.gov/pubmed/26131476

Coluzzi M, Sabatini A, Petrarca V, Di Deco MA (1979) Chromosomal differentiation and adaptation to human environments in the *Anopheles gambiae* complex. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 73, 483 – 497. [PubMed] https://www.ncbi.nlm.nih.gov/pubmed/394408.

Dellicour S, Gerard M, Prunier JG, Dewulf A, Kuhlmann M, Michez D 2017, Distribution and predictors of wing shape and size variability in three sister species of solitary bees. *PLOS* https://doi.org/10.1371/journal.pone.0173109.

Demari-Silva B, Suesdek L, Sallum MAM and Marrelli MT, 2014. Wing geometry of *Culex coronator* (Diptera: Culicidae) from South and Southeast Brazil. *Parasites & Vectors* 2014, 7:174. https://doi.org/10.1186/1756-3305-7-174.

Dujardin JP, 2011. Applications in Medical Entomology, Species Identification and Detection. *Genetics and Evolution of Infectious Diseases*, 16.6.1 DOI: 10.1016/B978-0-12-3890-1.00016-9, 2011 Elsevier Inc.

Ekgachai Jeratthitikul, Yago M, Hikida T 2013, Sexual dimorphism and intraspecific variation in wing size and shape of *Tongeia fischeri*. *Entomological Science*, Volume 17, Issue 3, pp 342-353. http://onlinelibrary.wiley.com/doi/10.1111/ens.12068/abstract.

Favia G, Lanfrancotti A, Spanos L, Sidén-Kiamos I, Louis C; 2001. Molecular characterization of ribosomal DNA polymorphisms discriminating among chromosomal forms of *Anopheles gambiae* s.s. *Insect Mol Biol*. 2001 Feb; 10(1):19-23. https://www.ncbi.nlm.nih.gov/pubmed/11240633

Gomez-Cendra PV, Paulin LE, Oroño L, Ovruski SM, Vilardi JC, 2016. Morphometric Differentiation Among *Anastrepha fraterculus* (Diptera: Tephritidae) Exploiting Sympatric Alternate Hosts. *Environmental Entomology*, Volume 45, Issue 2, pp 508–517, https://doi.org/10.1093/ee/nvv224. https://academic.oup.com/ee/article-abstract/45/2/508/2464994?redirectedFrom=fulltext.

Habtewold T, Duchateau L and Christophides GK 2016. Flow Cytometry analysis of the microbiota associated with the midguts of vector mosquitoes. Parasites and Vectors 2016, Vol: 9, ISSN: 1756-3305. http://dx.doi.org/10.1186/s13071-016-1438-0. https://parasitesandvectors.biomedcentral.com/articles/10.1186/s13071-016-1438-0

Hammond A, Galizi R, Kyrou K, Simoni A, Siniscalchi C, Katsanos D, Gribble M, Baker D, Marois E, Russell S, Burt A, Windbichler N, Crisanti A, Nolan T, 2016. A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae. Nature Biotechnology* **34**, 78–83 (2016) doi:10.1038/nbt.3439, https://www.ncbi.nlm.nih.gov/pubmed/26641531.

Harmston FC, Lawson EA. 1967. *Mosquitoes of Colorado*. Bureau of Disease Prevention and Environmental Control, U.S. Public Health Service. Atlanta. GA.

Hartog JD, Reijns R 2013. About I^3S. http://www.reijns.com/i3s/about/I3S_about.html

He-Ping Yang, Chun-Sen Ma, Hui Wen, Qing-Bin Zhan, and Xin-Li Wang, 2015. A tool for developing an automatic insect identification system based on wing outlines. *Sci Rep* 2015 5 12786 doi: 10.1038/srep12786. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4528224/

Jeratthitikul E, Yago M, Hikida T 2013, Sexual dimorphism and intraspecific variation in wing size and shape of *Tongeia fischeri Entomological Science*, Volume 17, Issue 3, pp 342-353. http://onlinelibrary.wiley.com/doi/10.1111/ens.12068/abstract.

Keeley SC, 1982. Morphological Variation and Species Recognition in the Neotropical Taxon *Vernonia arborescens* (Compositae), 1982. *Systemic Botany*, Vol. 7, No. 1, pp. 71-84. DOI: 10.2307/2418654 http://www.jstor.org/stable/2418654.

Lee Y, Marsden CD, Nieman C, Lanzaro GC, 2014. A new multiplex SNP genotyping assay for detecting hybridization and introgression between the M and S molecular forms of *Anopheles gambiae*. *Mol Ecol Resour*. 2014 Mar; 14(2):297-305. doi: 10.1111/1755-0998.12181. Epub 2013 Nov 11. https://www.ncbi.nlm.nih.gov/pubmed/24119184

Lorenz C Almeida F Fernanda Almeida-Lopes F Louise C, Pereira SN, Petersen V, Vidal PO, Virginio F, Suesdek L, 2017. Geometric morphometrics in mosquitoes: What has been measured? *Infection Genetics and Evolution*, Vol 54, 2017, pp 205-215. https://doi.org/10.1016/j.meegid.2017.06.029 .

Ngoagouni C, Kamgang B, Nakouné E, Paupy C and Kazanji M, 2015. Invasion of *Aedes albopictus* (Diptera: Culicidae) into central Africa: what consequences for emerging diseases? *Parasites & Vectors*2015 **8**:191 https://doi.org/10.1186/s13071-015-0808-3

Nosil P, Reimchen TE, 2005. Ecological opportunity and levels of morphological variance within freshwater stickleback populations. *Biological Journal of the Linnean Society*, Volume 86, Issue 3, 2005, Pages 297–308, https://doi.org/10.1111/j.1095-8312.2005.00517.x

Paz Garcia DA, Aldana-Moreno A, Cabral-Tena RA, García-De-León FJ, Hellberg ME, Balart EF, 2015. Morphological variation and different branch modularity across contrasting flow conditions in dominant *Pocilloporar*eef-building corals. *Oecologia*, Volume 178, Issue 1, pp 207–218. https://link.springer.com/article/10.1007/s00442-014-3199-9.

Pieterse W, Benítez HA, Addison P, 2017. The use of geometric morphometric analysis to illustrate the shape change induced by different fruit hosts on the wing shape of *Bactrocera dorsalis* and *Ceratitis capitata* (Diptera: Tephritidae). *Zoologischer Anzeiger* 269 110-116. https://doi.org/10.1016/j.jcz.2017.08.004. http://www.sciencedirect.com/science/article/pii/S0044523117300608

Powell JR and Tabachnick WJ, 2013. History of domestication and spread of *Aedes aegypti*, a review. *Mem Inst Oswaldo Cruz* 108 (Suppl 1) 11-17. Doi 10 1590/0074-0276130395, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4109175/.

Prudhomme J, Cassan C, Hide M, Toty C, Rahola N, Vergnes B, Dujardin JP, Alten B, Sereno D, Bañuls AL, 2016. Ecology and morphological variations in wings of *Phlebotomus ariasi* (Diptera: Psychodidae) in the region of Roquedur (Gard, France): a geometric morphometrics approach. *Parasit Vectors*, 9(1):578. DOI: 10.1186/s13071-016-1872-z; https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5109773/ .

Ranford-Cartwright LC, McGeechan S, Inch D, Smart G, Richterova L, Mwangi JM, 2016. Characterisation of Species and Diversity of *Anopheles gambiae* Keele

Colony**.** Published: December 29, *PLOS One* 2016 11(12) e0168999 doi 10 1371/journal pone 0168999 https://doi.org/10.1371/journal.pone.0168999

Sanzana MJ, Parra LE, Sepúlveda-Zúñiga E, Benitez HA, 2013. Latitudinal gradient effect on the wing geometry in *Auca coctei* Guérin (Lepidoptera, Nymphalidae). DOI: 10.1590/S0085-56262013005000045. *Revista Brasileira de Entomologia*: ISSN 0885- 5626. http://dx.doi.org/10.1590/S0085-56262013005000045

Soto IM, Carreira VP, Soto EM & Hasson E, 2008. Wing morphology and fluctuating asymmetry depend on the host plant in cactophilic *Drosophila*. *J. Evol. Biol*. 2 1 (2008) 598–609.

Soto, I.M., Soto, E.M., Corio, C., Carreira, V.P., Manfrin, M., Hasson, E., 2010. Male genital and wing morphology in the cactophilic sibling species *Drosophila gouveai* and *Drosophila antonietae* and their hybrids reared in different hostplants. *Environ. Entomol*. 39, 865–873. https://s3.amazonaws.com/academia.edu.documents/36712305/Soto_et_al_2010_Env_Entomol.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1514525893&Signature=OwxlDsy%2F1hILSvDzKWW4nDv13no%3D&response-content-disposition=inline%3B%20filename%3DMale_Genital_and_Wing_Morphology_in_the.pdf

Suman SD, Shrivasatava AR, Parashar BD, Pant SC, Agarwal OP, Prakash S, Variation in Morphology and morphometrics of eggs of *Culex quinquefasciatus* mosquitoes from different ecological regions of India, 2009. *J Vector Ecol* **34** (2) 191-9. doi: 10.1111/j.1948-7134.2009.00026.x.

Vidal PO, Suesdek L, 2012. Comparison of wing geometry data and genetic data for assessing the population structure of *Aedes aegypti*. *Infect Genet Evol.* **12** (3):591-6. doi: 10.1016/j.meegid.2011.11.013. Epub 2011 Dec 8. https://www.ncbi.nlm.nih.gov/pubmed/22178147

Vyas-Patel N, Mumford JD, 2017. I^3S Classic and Insect Species Identification of Diptera and Hymenoptera (Mosquitoes and Bees). doi: http://dx.doi.org/10.1101/090621. http://biorxiv.org/content/early/2017/02/11/090621

Vyas-Patel N, Ravela S, Mafra-Neto A, Mumford JD, 2015, Insect Wing Classification of Mosquitoes and Bees Using CO1 Image Recognition.
doi: https://doi.org/10.1101/034819
https://www.biorxiv.org/content/early/2015/12/23/034819.

Wilke ABB, Christe RO, Multini LC, Christe PO, Wil-da-Silva R, Carvalho GC, Marrelli MT, 2016 Morphometric Wing Characters as a Tool for Mosquito Identification. *PLOS* https://doi.org/10.1371/journal.pone.0161643.

Will WK and Rubinoff D 2004. Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics* 20 (2004) 47–55. Blackwell Publishing. file:///D:/Downloads/22.pdf.

Woodward CW, 1923, *The Wing Veins of Insects*. Reproduction paperback Publisher: Nabu Press (April 6, 2012) Paperback: 158 pages; Language: English

Wooton RJ, Functional Morphology of Insect Wings, 1992. *Annual Review of Entomology* Vol. 37:113-140
http://www.annualreviews.org/doi/10.1146/annurev.en.37.010192.000553

Yi Bai, Jia-Jia Dong, De-Long Guan, Juan-Ying Xie & Sheng-Quan Xu, 2016. Geographic variation in wing size and shape of the grasshopper *Trilophidia annulata* (Orthoptera: Oedipodidae): morphological trait variations follow an ecogeographical rule. *Nature Scientific Reports* 6, Article number: 32680 (2016), doi:10.1038/srep32680. https://www.nature.com/articles/srep32680