

SABCA data capture & processing methods

Compiled by Silvia Mecenero, Rene Navarro & Bennie Coetzer
(SABCA Project Coordinator)
March 2011

1. Data sources

Data for the atlas region (South Africa, Lesotho and Swaziland) were sourced from: 1) the digitization of existing museum and private collections, 2) field surveys by LepSoc members prioritized by gap areas in the atlas region and 3) the online virtual museum for the submission of photographic records from members of the public. Other data included academic research studies, field sightings, the literature and the first two South African Butterfly censuses.

Digitization of existing museum and private collections

Information on specimen labels in a collection was captured into a Microsoft Excel spreadsheet or, in cases where data were captured using software other than MS Excel, databases were extracted and converted into an MS Excel format. Spreadsheets in MS Excel were set up so that each column represented a field and each row represented one record. Where possible, specimens of the same taxon and collected on the same date and at the same locality by the same collector were grouped together as a series to form one record. Mandatory fields used were: Record accession number, Genus, Species, Subspecies, Locality description, Latitude, Longitude, Date, Collector, Number of specimens, Notes. Any other additional information was placed under relevant, optional field names, e.g. Altitude, Sex, Form, Drawer number, etc. If specimens did not have accession numbers, then these were assigned to them, e.g. initials of the collection owner and a number starting from one.

Locality information was captured exactly as it stood on the labels, to avoid the potential introduction of erroneous spelling corrections (e.g. both Ladysmith and Ladismith exist as different localities in separate provinces). Species names and dates were also captured as written on the labels. Where given, latitude and longitude coordinates were captured exactly as given on the labels to avoid errors in the conversion of coordinates to the DMS (Degrees-Minutes-Seconds) format.

Specimens were handled with great care to avoid breakage. Where necessary, labels were carefully removed from the specimen pin with a pair of tweezers in order to read the information on the label. If the handwriting was too small then a magnifying glass was used to improve legibility. Specimens were never removed from their pins.

Field surveys

LepSoc members were responsible for conducting the SABCA field surveys in South Africa. Surveys in each of the nine provinces in South Africa were coordinated by the LepSoc representative (Provincial Leader) for that province. Provincial Leaders were responsible for coordinating surveys within the relevant province. Data were collated in Lepidops, a programme developed by LepSoc over several years to support collection management and distribution data. The current version was developed around a relational database, Paradox, which provided a low cost environment for use by any LepSoc member. The version used for SABCA was V4.04.

Permits and licenses for the survey team were obtained from the provincial nature conservation agencies, various park authorities, South African National Parks, Department of Water Affairs and Forestry and some private forestry companies. Where necessary, permission was obtained from landowners or park/reserve managers prior to surveying an area.

Surveys were guided by a field protocol. Generally, a survey locality was an area with a radius of about 100m, and a GPS was used to determine the latitude and longitude coordinates of the centre point of each locality (Degrees, Minutes and Seconds). All butterfly species at a survey locality were recorded. It wasn't necessary to sample all species as some were identifiable without sampling. A maximum number of six pairs of butterflies per species, per locality, per survey, was allowed for collection by each field survey team, unless otherwise stated on the relevant permit. Collecting restrictions on Red Listed species, as laid out in LepSoc's Code of Conduct (www.lepsoc.org.za) were followed. For small colonies of butterflies, the number collected was determined in the field but this was always lower than the amount allowed by the permit. Woodhall's (2005) field identification guide was used as a standard reference for identifying species. Apart from any additional permit conditions, holotype specimens or new species were deposited in a recognised South African museum or scientific institution, preferably the Ditsong National Museum of Natural History (formerly the Transvaal Museum). Team members were able to keep any voucher specimens surplus to those needed to meet

permit requirements within their private collections. Additional surplus voucher specimens were sent to the Ditsong Museum for safe-keeping.

Three field seasons were conducted during SABCA, each beginning in August and ending the following May. Field surveys were initiated in August 2007. The first and second field seasons were guided by a gap analysis based on the expert knowledge of LepSoc members G.A. Henning and D.A. Edge as well as existing data in LepiDops, at a one degree grid square scale. The third season was guided by gaps in the coverage map based on data in the SABCA database at the beginning of this season, at a QDGS scale.

Online virtual museum

The SABCA VM is accessible at <http://vmus.adu.org.za>. It was developed at the ADU using open-source software: MySQL provided the database backend and the front end was written in the general-purpose scripting language, PHP.

Initially, photographs of any life stages of butterflies were submitted by members of the public via email, along with the following basic information: observer name, coordinates latitude, coordinates longitude, locality, closest town, province, country, year, month, day, number of photos, life stage, notes. Photographs of both the upper and under sides of the adult wings were preferable to facilitate identification. Locality information was verified (provided coordinates were checked that they reflected the locality descriptions and if coordinates were missing then they were determined based on the locality descriptions, to Degrees and Minutes) and photos were cropped prior to uploading photographic records onto the online VM. In the final year of SABCA photographic submissions were no longer accepted via email, but via an online submission procedure on the VM site which captured the same basic information that was required for the email submissions as well as the photographs which required an extra confirmatory step before becoming visible on the VM site. Photographic specimens were identified online by a panel of experts. For common and easily identifiable species, identifications could be confirmed by one panel member only. For species which were difficult to identify, at least two panel members had to agree on the identification before it was confirmed.

Other data, including the butterfly censuses

Field sightings were received from LepSoc and non-LepSoc members. Peer-reviewed journals and LepSoc's Metamorphosis journal contributed to data from the literature. South Africa's first two Butterfly Censuses took place over the weekend of 24-25 April 2010 and the week of 9-17

October 2010, and were open to both beginners and experts. Each participating team registered a locality, preferably no larger than about 1km in radius, and a minimum of six hours were spent in the registered locality observing and counting butterflies (e.g. sabca.adu.org.za/bcw.php for more details on the census protocol). Teams made use of data capture forms to note their observations and census details (e.g. sabca.adu.org.za/bcw_dataform.php). Pamphlets of the most common species in each province were made available online to assist with the identifications (e.g. sabca.adu.org.za/bcw_pamphlets.php), otherwise photos of the butterflies were submitted to the VM for identification by the expert panel.

2. Data processing and vetting

Formatting

Data were formatted into a standard format in MS Excel using the following fields:

Accession number (if a record did not have an accession number, then in most cases one was assigned during the digitization process)

Collection name (institution, collector or observer name)

Genus, Subspecies, Species (see below)

Locality (full locality description as appeared on the specimen label or as provided)

Closest town (all records were assigned a closest town by the project data technicians, based on the locality description; these values did not appear on specimen labels unless a specimen's locality description was indeed a town)

Province (determined by the project data technicians and/or updated to reflect the current nine provinces in South Africa)

Country (updated to reflect the current country names)

Latitude (South coordinate)

Longitude (East coordinate)

Year, Month, Day (Date of record)

Observer (Name of the collector or observer)

Record type (Voucher = collected specimens, Bred = specimens that were collected and then bred out, Photo = photographic specimen, Observer_U = observation made by a non butterfly expert, Observer_V = observation made by a butterfly expert)

Number of specimens (the number of specimens within the record)

Accuracy code (see below)

Coordinate source (see below)

Altitude (optional field)

Notes (any additional notes attached to specimens or observations)

Taxonomy

Taxonomic updates relied on the input of LepSoc's Conservation Advisor (G.A. Henning), as well as the assessment authors, who based the updates on the geographic locality of specimens as well as on knowledge of a taxon's taxonomic history. In some cases, LepSoc members were able to check and confirm specimen identifications at various museums and institutions, but specimens were generally not examined to confirm their taxonomic updates. In some cases, specimens that appeared to be possible new taxa were allocated to the closest existing taxon and flagged for further investigation.

Geo-referencing

Where latitude and longitude coordinates were missing, data were geo-referenced to Degrees and Minutes, provided that localities were adequately described. If coordinates were already provided, they were checked to ensure that they accurately reflected the locality descriptions. If coordinates were given as decimal degrees or decimal minutes, these were converted to Degrees and Minutes or Degrees, Minutes and Seconds.

Accuracy and source codes were assigned to the coordinates for each record. The accuracy codes used were: 7 (over 1000m), 17 (centroid of QDGS), 20 (centroid of farm), 21 (centroid of reserve/park), 22 (centroid of urban area), 23 (centroid of locality), 26 (integer minute), 27 (decimal minute), 28 (integer second) and 29 (decimal second). The source codes used were: Map (e.g. MapSource version 6.5, Garmin Ltd., 1:50 000 maps, road atlases), Planet (online geographical database, www.mapplanet.com), Geonames (online geographical database, www.geonames.org), Gazetteer (compiled by the project as localities were geo-referenced and from geo-referenced localities from collectors), Google (as a last resort, an internet search was made using the Google search engine), Google Earth software (version 5.1), GPS-act (for actual latitude and longitude coordinates determined using a GPS), and GPS_approx (for approximate latitude and longitude coordinates determined using a GPS).

Data upload

All processed data were uploaded into a single comprehensive database using MySQL. MySQL is open source software and runs on more than 20 platforms, including Linux, Windows, OS/X and Netware, and is therefore fully portable to most other modern databases. Prior to upload

into the database, each data record was assigned a unique SABCA record number. During the upload process, each record was assigned a locus (i.e. QDGS). The entire database is publicly accessible via SANBI's online data portal (SIBIS:SABIF data portal; <http://sibis.sanbi.org/>).

Vetting during the conservation assessment process

During the online compilation of the conservation assessments, assessment authors checked the online distribution maps for outlier records. Outlier records were flagged as Questionable and were excluded from the assessments. Some records were also flagged as Historical, indicating records of localities which no longer exist due to habitat degradation. All other records were flagged as Accepted. Only records submitted up until the assessment cut-off date (20 July 2010) were checked (these records were given an extra flag of 1 in the field "Asmt_flag" of the distribution database). Records submitted after the assessment cut-off date have not been checked for Questionable or Historical records, they are all flagged as Accepted as well as 0 in the field "Asmt_flag" of the distribution database).

3. Data limitations

- Most of the older data had to be geo-referenced to degrees and minutes, based on the locality description, as actual points of observation were unknown for these.
- Records reflect mainly the presence of taxa within QDGSs, and not their absence.
- The collection or observation of specimens was usually biased towards common, easily observed, easily photographed or interest-group taxa, and thus records are taxonomically biased.
- Many records are centered around urban areas or areas within easy reach, thus geographical coverage was incomplete, with most gaps in the Northern Cape, North West, Free State, Lesotho and Swaziland. Thus records are geographically biased.
- Field surveys were not standardised and no measure of effort could be made.
- For all records there is no measurement of observer effort, thus precluding assessments of bias and influencing data analyses.