Harnessing social media data to track species range shifts

Shawan Chowdhury*[1,2,3,4,5], Niloy Hawladar[6], Ripon C Roy[6], César Capinha[7,8], Phillip Cassey[5], Ricardo A. Correia[9,10,11], Gideon Gywa Deme[12,13], Moreno Di Marco[14], Enrico Di Minin[9,10,15], Ivan Jarić[16,17], Richard J. Ladle[19], Jonathan Lenoir[20], Mohammad Momeny[9], Jooel J Rinne[9], Uri Roll[21], Aletta Bonn[1,2,3]

* Corresponding author (dr.shawanchowdhury@gmail.com)

[1]Institute of Biodiversity, Friedrich Schiller University Jena, Jena, Germany
[2]Department of Biodiversity and People, Helmholtz Centre for Environmental Research - UFZ, Leipzig, Germany
[3]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany
[4]Czech University of Life Sciences Prague, Faculty of Environmental Sciences, Prague, Czech Republic
[5]School of Biological Sciences, The University of Adelaide, Adelaide, SA, Australia
[6]Department of Zoology, University of Dhaka, Dhaka, Bangladesh
[7]Centre of Geographical Studies, Institute of Geography and Spatial Planning of the University of Lisbon, Lisbon, Portugal
[8]Associate Laboratory Terra, Lisbon, Portugal
[9]Department of Geosciences and Geography, University of Helsinki, Helsinki, Finland
[10]Helsinki Institute of Sustainability Science, University of Helsinki, Helsinki, Finland
[11]Biodiversity Unit, University of Turku, Turku, Finland
[12]Department of Biology, Case Western Reserve University, Cleveland, Ohio, USA
[13]Department of Science Laboratory Technology, University of Jos, PMB 2084, Jos, Nigeria
[14]Department of Biology and Biotechnologies Charles Darwin, Sapienza University of Rome, Rome, Italy
[15]School of Life Sciences, University of KwaZulu-Natal, Durban, South Africa
[16]Université Paris-Saclay, CNRS, AgroParisTech, Ecologie Systématique Evolution, Gif-sur-Yvette, France
[17]Biology Centre of the Czech Academy of Sciences, Institute of Hydrobiology, České Budějovice, Czech Republic
[19]Institute of Biological and Health Sciences, Federal University of Alagoas, Maceió, Brazil
[20]UMR CNRS 7058, Ecologie et Dynamique des Systèmes Anthropisés (EDYSAN), Université de Picardie Jules Verne, Amiens, France
[21]Mitrani Department of Desert Ecology, The Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Midreshet Ben-Gurion, Israel

Abstract

Biodiversity monitoring programs and citizen science data remain heavily biased towards the Global North. Incorporating social media data can complement existing gaps, especially in megadiverse countries with limited records, but whether such data can significantly improve our understanding of range-shifting species is unknown. Here, we collated locality data from Flickr and Facebook, in addition to occurrence data from the Global Biodiversity Information Facility (GBIF). We tested whether social media data improved our knowledge of the range dynamics of a rapid range-shifting butterfly, tawny coster (*Acraea terpsicore*), when compared to GBIF-only data. Social media data increased occurrence records by 35%. The proportion of social media data was higher in countries poorly represented on GBIF; however, we also obtained new distributional information from well-represented countries (e.g., Australia and Malaysia). We constructed ecological niche models (ENM) with data from both sources. ENMs based on the social media data highlighted greater expansion rates to higher latitudes and elevations compared to ENMs based on GBIF data only. Our results highlight the potential of harnessing social media data to track biodiversity redistribution in response to climate change.

Main

The current era is marked by a pressing biodiversity crisis (Dirzo et al., 2014; Pimm et al., 2014; Diaz et al., 2019), driven by various factors such as agricultural intensification and expansion, habitat loss and fragmentation, biological invasions, and climate change (Wilson, 1989; Butchart et al., 2010; Joppa et al., 2016; Maxwell et al., 2016; Jaureguiberry et al., 2022). In response to these combined pressures on the environment, many species have shifted their distributions (Chen et al., 2011; Yackulic et al., 2011; Lenoir & Svenning, 2015; Lenoir et al., 2020; Chan et al., 2024). While range-shifting can be feasible for migratory and highly mobile species, which can easily expand towards new areas (Chowdhury et al., 2021a, b), it poses a significant challenge to narrow-range and low-mobility species that may be unable to reach new suitable habitats, and thus may be doomed to range contraction and ultimately extinction (Pound et al., 2006; Freeman et al., 2018; Chowdhury, 2023). Many taxonomic groups contain species that have recently shifted range size, including insects (McCain & Garfinkel, 2021), mammals (Santos et al., 2017), birds (Rushing et al., 2020), amphibians (Nowakowski et al., 2017), plants (Auffret & Svenning, 2022; Iseli et al. 2023), and marine organisms (Poloczanska et al., 2016). When a species migrates to a new habitat, it may be restricted to similar climatic conditions (niche-conserving species; Wiens et al., 2010) or adapt to different ones (niche-shifting species; Guisan et al., 2014; Di Marco et al., 2021). For instance, a native range shifter may track shifting isotherms altitudinally or latitudinally to remain within the same climatic space, while a non-native invader may spread from its introduction point to exploit new climatic spaces. The colonisation or extirpation rates of a range-shifting species (being native or non-native) depends largely on prevailing landscape conditions: if only a small amount of landscape is suitable, the establishment of a founding population might be uncertain or slow, or the founding population may go extinct due to Allee effects (Kuussaari et al., 1998; Hodgson et al., 2012; Blackburn et al., 2016).

The issue of biodiversity redistribution in response to human activities is a global concern (Lenoir & Svenning, 2015; Lenoir et al., 2020). However, our understanding of the issue is limited and biased, both geographically and taxonomically (Lenoir & Svenning, 2015; Feeley et al., 2017). For instance, a systematic review of 258 peer-reviewed studies reporting species range shifts revealed that nearly 40%

of the identified range shifts (12,415 species) were from flowering plants, and 22%, 4%, 12% and 0.5% were from insects, fish, birds and mammals, respectively (Lenoir et al., 2020). For plants, the data is strongly biased towards flowering plants with very few records for ferns, mosses, etc. while for animals, the data is strongly biased towards birds and fish. Likewise, in insects, there is a strong bias towards specific groups, including orthoptera and dragonflies (Lenoir et al., 2020). This taxonomic bias is compounded by a severe geographical bias, with the majority of studies coming from countries of the Global North and only a limited number from the tropics and the Global South (Lenoir et al., 2020; Lawlor et al., 2024; Parker et al., 2024), including many megadiverse countries. This supports the claim that researchers' efforts to document species range shifts have been heavily biased towards certain regions and taxa, meaning none of the 'global reviews' are truly global (Feeley et al., 2017). Assessing species range shifts requires detailed species distribution data, typically unavailable for species from the Global South (Hortal et al., 2015; Hughes et al., 2021). This, in turn, biases our global overview of species undergoing climate-induced range shifts. To improve our understanding of species redistribution at a global scale, it is essential that we compile data from all available sources to better test how species are responding to global change drivers.

With the rapid technological advances involving mobile phones, digital cameras, and fast internet has revolutionised data collection (Van Klink et al., 2022; Sheard et al., 2024), but also anyone from anywhere in the world can now share their biodiversity observations on a variety of digital platforms (e.g., apps, web pages, and social media; Chandler et al., 2017; Pocock et al., 2018; Toivonen et al., 2019; Caley & Cassey, 2023). If appropriately harvested, such data can be crucial in filling knowledge gaps in biodiversity distribution and monitoring and thus help us answer questions about how species respond to global changes (Jarić et al., 2020; Soriano-Redondo et al., 2024). This is especially relevant for countries where systematic biodiversity monitoring programs are uncommon and which are poorly represented in global biodiversity repositories (Chowdhury et al., 2023a; Marcenò et al., 2021; Mota et al., 2022). Indeed, recent studies have demonstrated that biodiversity data posted on Facebook can sometimes be much more comprehensive than the data available from the Global Biodiversity Information Facility (GBIF) (Chowdhury et al., 2023a). Using a comprehensive set of biodiversity data in ecological research is key in assessing species redistribution in the context of global change. For example, combining social media data with GBIF data can substantially improve conservation assessments (Chowdhury et al., 2024a), which can be useful in identifying priority conservation areas (Chowdhury et al., 2024b). Nevertheless, biodiversity observation data from social media have rarely been used in conservation assessment studies (Di Minin et al., 2015) and for tracking range-shifting species (Sbragaglia et al., 2024). Here, as a proof of concept, we aim to address this important knowledge gap by harnessing social media data to evaluate the distribution pattern of an ecologically important species that is undergoing a rapid range expansion in response to global changes: the tawny coster butterfly (*Acraea terpsicore*).

We compiled the distribution data of the tawny coster butterfly from social media (Facebook and Flickr), and combined these with GBIF records to assess the extent to which social media data adds new locality information where it occurs. We further analysed the data into two groups: GBIF-only and combined (GBIF and social media) data. We organised the data in five time intervals of four years each and fitted ecological niche models to identify if the geographic range distribution of this species would differ from the known range extent (Chowdhury et al., 2021a) and how this would impact the calculation of range expansion rates. Finally, we ran a niche assessment analysis to identify whether the additional social media data identified novel combinations of the environmental niche space. Building on our findings, we provide recommendations on how to use social media data to answer general ecological questions relating to species' biogeography and macroecology.

Results

Our cleaned and compiled dataset included 6459 occurrence records, of which 65% (4206) were from GBIF and 35% (2253) were from social media (Flickr: 5%, Facebook: 30%). We noticed marked differences between data sources when we compared the number of occurrence records across countries (Figure 1). For most countries, the number of occurrence records substantially increased after including social media data (e.g., data increased from 10 to 224 for Bangladesh, 262 to 468 for Malaysia; Figure 1A, B). The distribution of the tawny coster butterfly is known from 17 countries (Chowdhury et al., 2021a), and we obtained a higher percentage of data than GBIF from social media for five of these countries (Figure 1A, B; Supplementary Data S1). In countries with the most occurrence records, the percentage of the data coming from social media was generally lower, but >10% in all cases (range 10-44%). For example, we obtained 3096 occurrence records from India, of which 64% (1968) were from GBIF, and 36% (1128) were from social media (Figure 1B). The percentages of species occurrence records from social media were higher for countries with lower number of total occurrence records (Figure 1B).

When we analysed the temporal distribution of occurrence records, we noticed substantial differences over the years (Figure 1C). While the initial period, from 2005 to 2007, had a larger percentage of occurrence records from GBIF, subsequent years (2008-2018) were characterized by a higher percentage of occurrence records from social media, except for 2013. Following a substantial decline during 2017-2022, the proportion of social media data stabilized recently (Figure 1C).

The addition of social media data in ecological niche models contributed substantially to the identification of potentially new suitable areas for the tawny coster. The total surface area predicted to be suitable for the tawny coster is in general larger when combining social media data with GBIF data than when relying on GBIF-data only, at least during the periods 2005-2008, 2017-2020 and 2021-2024 (Figure 2). For the period 2009-2012, we found the opposite pattern. New suitable areas identified with the addition of social media data were mostly distributed in South Asia (especially towards higher elevations in the Indo-Himalaya region) during 2005-2020, while new suitable areas identified with the addition of social media data during 2021-2024 were distributed throughout the entire region and especially towards higher latitudes (Figure 2). When we combined all the suitability maps over the five time intervals, we found that predictions from the models relying on GBIF-only data missed many areas at higher latitudes and at mid- to high elevations (Figure 2G). In terms of range expansion, the combined data captured a larger expansion area initially; however, it slightly declined afterwards and increased again recently (Figure 2H).

To understand the benefit of adding social media data to fit our models, from the perspective of the environmental niche space, we conducted a Principal Component Analysis (PCA) at each of the five studied time intervals separately for both GBIF and combined data. The PC1 and PC2 axes explained from 59-63% of the total variance. Across all 5 consecutive time periods, the model combining GBIF data with social media data captured a broader niche space than the model relying on GBIF-only data (Figure

3). While the niche overlap between the model relying on GBIF-only data and the model combining social media data with GBIF data was fairly large, reaching 65% and 76% during 2005-2008 and 2009-2012, respectively, the overlap was much smaller during the three subsequent periods (49%, 37%, and 39% during 2013-2016, 2017-2020, and 2021-2024, respectively). The PCA identified precipitation (monthly total), maximum temperature, and elevation to be the most important environmental variables determining the differences in the covered environmental niche space between the model relying on GBIF-only data and the model combining both GBIF and social media data. Across time intervals, the GBIF-only data failed to capture regions with lower maximum temperature, lower precipitation (monthly total), and higher elevation.

Discussion

Social media data can help reduce the global biodiversity data shortfall (Di Minin et al., 2015; Chowdhury et al., 2023a) and improve our understanding of biodiversity (re)distribution for conservation purposes (Chowdhury et al., 2024a). Yet, such data have rarely been used in large-scale studies (Di Minin et al., 2015). Using standardised protocols to scrape data from social media, we harvested more occurrence data for modelling the potential redistribution of a rapidly expanding species than would typically be used in traditional distribution modelling studies that rely on GBIF data only. We demonstrated that social media data can identify new distribution: occurrence records increased by 53% (4206 to 6459). For at least some of the recent years, the percentages of social media data decreased slightly, which could be explained by COVID pandemic and lockdowns (Chowdhury et al., 2024b), people travelled less and reduced tourism could have resulted in less biodiversity records posted. When combining occurrence records from social media with those from GBIF and fitting models to project habitat suitability maps, the suitable area of potential occupancy increased for three of the five studied time intervals, and the total amount of additional areas increased over time compared to a model relying on GBIF data only.

Systematic biodiversity monitoring programs and citizen science provide important biodiversity data sources for scientists and conservation biologists (Mesaglio & Callaghan, 2021). In addition, social media data has a key potential to improve our basic understanding of species' distribution and spread, even in better surveyed (e.g., developed) countries. For example, biodiversity data from Australia is comparatively well-represented in global biodiversity repositories like GBIF, but we still obtained many new localities from social media that represented uncharted conditions, from a GBIF perspective, within the climatic space. The total number of occurrence records increased by 12% (440 to 493) and the suitable area of potential occupancy increased by 9% (1.64 million km^2 to 1.79 million km^2). Such gains were even more pronounced for meagdiverse countries of the Global South, such as Bangladesh, where the total number of occurrence records retrieved from social media was 22.4 times higher than from GBIF. This illustrates the enormous potential of social media data to reduce the global biodiversity data shortfall as a means to better track range-shifting species. By doing so, we managed to identify many new localities (at higher latitudes and at higher elevations, chiefly representing climate conditions from colder environments with lower maximum temperature and lower precipitation) that the tawny coster butterfly might colonise in the future in response to climate warming. We also showed that social media data helped to capture a broader niche space exploited by the tawny coster butterfly, some of which were not captured by the model relying on GBIF-only data.

We used two social media channels – Facebook and Flickr – to harvest more occurrence data for the tawny coster butterfly. While the additional data substantially improved the performances of our species distribution model and niche assessments, we faced several obstacles. First, we used machine learning to automatically scrape data from Flickr (following Hausmann et al., 2018) while we had to manually extract data from Facebook (following Chowdhury et al., 2024b), which was a time-consuming task compared to an approach that relies solely on artificial intelligence (AI). In the future, it should be possible to develop an automated approach to extract species' occurrences from Facebook, which would save a substantial amount of time (Jarić et al., 2020; Correia et al., 2021; Chowdhury et al., 2024b). For example, Castro et al. (2024) showed that the success rate of AI models in extracting information from unstructured text is quite high, making them valuable tools for managing ecological data efficiently. Second, we faced two major data issues when using Flickr: photographs with no location data and photographs erroneously flagged as the tawny coster. Because of that, we could only use 5% of the data we initially scraped from Flickr. To handle this issue, it is important to improve the Flickr data extraction process, by carefully checking individual photographs, and validating whether or not they represent what we are looking for. Finally, many photographs shared on social media might not be the species the photographers assume to be. To handle this issue, having someone in the group with taxonomic expertise is essential, especially the people who are extracting records from social media and validating species information.

It is important to think strategically to get the maximum value from social media data. We only used Facebook and Flickr as social media platforms, which tend to be less popular in some countries. Including other popular platforms for particular countries (e.g., Weibo in China or possibly Instagram in other countries) could provide many more new records. We recommend future studies assessing data quality performance across several social media channels. Furthermore, we had to remove many records due to data quality issues (e.g., the locations were unspecified, and photographs were unclear). To solve these issues, group moderators are needed and should maintain strict rules about sharing biodiversity observations so that everyone knows the species' details.

Although open data would revolutionise scientific research, it is important to think differently in the case of threatened species, as such data can increase threats (e.g., poaching, disturbance; Bergman et al., 2022; Di Minin et al., 2015, 2022). Group moderators and regional legal authorities should deal with such issues. When using social media data to extract biodiversity data, personal information should be carefully handled and potential intentional and unintentional physical and mental harm to the photographers should also be carefully considered (Di Minin et al., 2021). To mitigate these risks and ensure user safety, adequate practices such as data minimisation, anonymisation, and strict data management protocols should be adopted (Di Minin et al., 2021).

The Kunming-Montreal Global Biodiversity Framework (CBD, 2022) aims to ensure the best data available for conservation assessments (Target 21), and to protect 30% of the Earth by 2030 (Target 3). Here, by comparing data distribution between the most comprehensive global biodiversity repository (GBIF) and social media, we showed that biodiversity data shared on social media can improve scientific knowledge on species distributions, even in countries that are well represented in global biodiversity repositories like GBIF or iNaturalist. Due to environmental changes, range-shifting species (including invasive species) are expanding rapidly, and social media data are especially powerful in this situation as they allow for almost real-time monitoring, which is not typically possible when relying on GBIF data solely. This makes social media data especially useful to set up 'early warning' systems of species colonisation (Soriano-Redondo et al., 2024). There is potentially even more data available if we develop more powerful digital tools. In addition to other platforms that are currently difficult to access (e.g., Instagram), there is also incidental (or secondary) biodiversity data (e.g., posted photographs of flowers that, by chance, have a butterfly on them; Pernat et al., 2024). These data might come into play with improvements in automated species recognition tools. Such approaches can help better understand and track ongoing species' movements and future biological invasions (Capinha et al., 2024; Cardoso et al., 2024). The current conservation literature is highly biased, chiefly stemming from North America and Western Europe (Di Marco et al., 2017; Dawson et al., 2024), because of significant and long-standing human capacity limitations in the tropics. Our findings suggest that combining data from multiple sources can eventually help answer key ecological questions, especially for countries with limited biodiversity observation records currently registered in global biodiversity repositories.

Methods

Tawny coster

The tawny coster has a well-documented geographic range area (Chowdhury et al., 2021a), and its charismatic status, like many butterfly species, attracts high public attention on social media. This butterfly is native to the Indian subcontinent (India, Bangladesh, and Sri Lanka) (Braby et al., 2014; Chowdhury et al., 2021a) and since the 1980s it has rapidly expanded its range to other parts of South Asia (e.g., Bhutan, Nepal, and Pakistan) and Southeast Asia (e.g., Malaysia, Singapore, and Thailand), eventually entering Australia in April 2012 (Braby et al., 2014). It was first recorded in Australia near Darwin in the Northern Territory (Sanderson et al., 2012). In subsequent years, the species started to spread towards Western Australia (till Kimberly), before the spread abruptly shifted towards Queensland. Since its arrival in Australia, the tawny coster has expanded within the country at a rate of approximately 135 km/year, while remaining within its native climatic niche (Chowdhury et al., 2021a). The species can cover a wide range of habitats and can migrate long distances (Chowdhury et al., 2021b), which might have facilitated its documented expansion.

Data

We collated occurrence records and locality data for tawny coster’s sightings from three different sources: GBIF, Flickr, and Facebook. For all the data sources, we selected data from January 2005 to May 2024 to maintain a comparable sampling period. We removed potential duplicates in occurrence records using two approaches: (I) exact duplicates (i.e., rows containing the exact same values for all the columns), that we used to compare the distribution of occurrence data, and (II) occurrence records falling within the same 4.65 × 4.65 km^2 grid cells (only using longitude and latitude coordinates), which we used for niche modelling and the niche overlap analysis.

We downloaded GBIF data manually from the website (https://www.gbif.org/; GBIF, 2024). The GBIF portal is a collection of hundreds of citizen science applications (Heberling et al., 2021), so we did not look for citizen science data from other potential sources (e.g., iNaturalist). When downloading occurrence records from GBIF for the tawny coster, we kept only the presence data with coordinate uncertainty below 10 km. It should be noted that many GBIF occurrence records lack information on coordinate uncertainty, and our choice to ignore these records resulted in a reduced (but more reliable) sample.

We used a Python (https://www.python.org/) script, which uses the Flickr's application programming interface (API) (https://www.flickr.com/services/developer) and its keyword search, to collect all publicly available Flickr posts related to the tawny coster. We used the scientific name and the English common name of the species as a set of keywords for the searches. We then deduplicated the data and removed any posts not containing a geotag. Using the URLs of the posts, we manually double-checked all the photographs of the tawny coster.

For Facebook data, we followed the protocol developed by Chowdhury et al. (2024b). Specifically, the entire data extraction process was divided into three steps. First, we searched for butterfly groups using a combination of taxon and country names (Supplementary Table S1). Here, we collected all the known distribution (17 country names used as keywords) of the tawny coster from Chowdhury et al. (2021a). When searching for Facebook groups, we included ten more countries from the surrounding area. With these 27 keywords corresponding to 27 countries, we identified 41 Facebook groups from 17 countries (Supplementary Table S1). Second, in each Facebook group, we searched twice using both the scientific name (*Acraea terpsicore*) and the common name (tawny coster). We carefully went through each photograph and validated the species' information. From each photograph, we extracted date (day, month, year), location, and photographer's information. We excluded photographs if their quality was unsuitable for identification up to the species level, if a specific date and location were not provided, and if the location provided in the photographs was unspecific (> 10 km uncertainty). Finally, we used Google Maps (https://www.google.com/maps) to georeference the location information and get the longitude and latitude coordinates.

During the initial data cleaning process, we removed all duplicate records (same coordinates) and only kept records between January 2005 and May 2024. Our compiled dataset included 6459 occurrence records (GBIF: 4206; Flickr: 325; Facebook: 1928). We provide the Facebook and Flickr data in the supplementary material (Supplementary Data S1), while the GBIF data is publicly available (GBIF, 2024).

We used the TerraClimate database (https://www.climatologylab.org/terraclimate.html; Abatzoglou et al., 2018) to obtain climatic predictor variables at a yearly resolution (2005-2023) at 21.625 km^2 resolution. The climatic data for 2024 is yet to be published. We downloaded ten climatic predictor variables from TerraClimate: actual evapotranspiration, climate water deficit, potential evapotranspiration, precipitation (monthly total), soil moisture, maximum temperature, minimum temperature, wind speed, and the Palmer Drought Severity Index. We downloaded the elevation data from the WorldClim (https://www.worldclim.org/data/worldclim21.html) database at the same resolution, which corresponds to 21.625 km^2 resolution at the equator (4.65 km × 4.65 km).

Data preparation

We analysed range-shift dynamics of the tawny coster by splitting the 2005-2024 period into five intervals of four years each (2005-2008; 2009-2012; 2013-2016; 2017-2020; and 2021-2024) and assigned the occurrence and climatic data, available at a yearly resolution, accordingly to each of those 5 periods. For the occurrence data, separately for the GBIF and social media datasets, we grouped the observations into year intervals and kept a single occurrence record per grid cell if several occurrences from subsequent years of the same period were available at 21.625 km^2. Afterwards, we compared the number of occurrence records between the GBIF-only dataset and the social media dataset and kept the exact same number of records between both datasets, using a randomized sampling procedure in R. For example, if the GBIF and social media data contained X and Y occurrences, respectively, for a given period t, with $X > Y$ at t, then we subsampled X into a smaller subset x such that $x = Y$. By doing so, we managed to balance the sampling effort between the GBIF and social media data, thus limiting any improvement in model accuracy between both datasets that would be due to sample size. Finally, we merged the GBIF and social media datasets to create the combined dataset, so we had two datasets for the subsequent analyses GBIF-only and combined datasets.

For the environmental data, we cropped the layers to the study extent (xmin = 60.875, xmax = 158.9583, ymin = -54.75, ymax = 53.54167) and calculated the mean climatic conditions over the four years of each time period and did that for each of the ten climatic variables (there was no yearly elevation data). Given that the climatic data for 2024 is unpublished, for the last time period (2021-2024) we computed mean climatic conditions over three years, instead of four years. We checked for multicollinearity issues among the predictor variables and removed highly correlated ones ($|r| > 0.75$). Consequently, we removed four variables and kept seven remaining variables for the final analysis: climate water deficit, precipitation (monthly total), soil moisture, maximum temperature, wind speed, Palmer Drought Severity Index, and elevation.

Habitat suitability maps

To obtain habitat suitability maps for the tawny coster butterfly, we fitted MaxEnt species distribution models (SDMs) (Elith et al., 2010; Phillips et al., 2017) in R (R Core Team, 2024), using the ENMeval package (version 2.0.4; Kass et al., 2021). We ran the model twice for each of the five periods, once for the GBIF-only dataset and a second time using the combined datasets.

We fitted SDMs using the following settings: seven predictor variables (i.e., the ones selected after removing highly correlated ones, see the previous section entitled ‘Data preparation’), and 10,000 randomly generated background points at 21.625 km^2 resolution (2.5 arc minute). For all five time periods and for both the model relying on GBIF-data only and the model relying on the combined dataset, we used the exact same set of pseudo-absences by randomly selecting 10,000 background points across the entire study extent (Supplementary Figure S2). We did that to avoid the background selection strategy to affect the model outputs when comparing model performances over time and between data sources (GBIF-only vs. combined data). Before fitting the model, we removed duplicate values in each raster pixel and created a 500 km buffer around the spatial records. We cropped the environmental variables to the buffered region to limit model overfitting. We assigned the records to grid cells and then randomly assigned grid cells to particular folds (Kass et al., 2021). We used the ‘checkerboard2’ evaluation method (with the presence and background points), which handles overinflation of model performance, at least from biased sampling. This evaluation method partitions geospatial records and background points into evaluation bins to reduce spatial autocorrelation between points in the testing and training bins. To improve MaxEnt’s modelling performances, we performed a calibration procedure by fitting the model under different combinations of parameters and hyperparameters. Specifically, we fitted the model under six feature class combinations (L, LQ, H, LQH, LQHP, and LQHPT, where L is linear, Q is quadratic, H is hinge, P is product, and T is threshold) and eight different regularisation multipliers (0.5 to 4 at 0.5 intervals). While the feature class allows MaxEnt to develop composite models to ensure a good fit to the data, regularisation multiplier values control model overfitting (Kass et al., 2021).

Overall, there were 48 models (6 (feature class) x 8 (regularisation multiplier) for each data group (GBIF-only vs. combined data) in each time interval. We chose the best model with the lowest Akaike Information Criterion (AICc) (Kass et al., 2021). Using the raster package in R (Hijmans, 2023). We used the 10% omission rate threshold value and transformed the suitability map into binary classes based on the threshold value (1 for presence with suitability value ≥ threshold value; and 0 for absence with suitability value < threshold value). We calculated the centroid position of these binary maps and used the geosphere (Hijmans, 2022) package in R to calculate the range expansion rate.

Niche assessment

We used the ecospat R package (Broennimann et al., 2023) to evaluate, for each time period separately, whether the use of additional data from social media led to significant differences in the realised niche space occupied by the species. We quantified the niche overlap between the GBIF-only and combined dataset, for each of the five time periods separately. We used the same seven environmental variables that were used earlier for fitting SDMs. We extracted the environmental data corresponding to the occurrence records and ran the Principal Component Analysis (PCA) to reduce dimensionality. We transformed the first two components (PC1 and PC2) into density by kernel smoothers. The PC1 and PC2 explained from 59-63% of the total variance. We quantified niche overlap using Schoener's D metric and assessed the statistical significance through niche equivalency and similarity tests following the methods described by Warren et al. (2008) and Broennimann et al. (2012). These tests were implemented to determine whether the observed niche overlap was greater than expected by chance, providing insight into niche dynamics over time.

Data and code availability

We have attached the social media data in the supplementary section (Supplementary Data S1) and the GBIF data are publicly available (GBIF, 2024).

All the R scripts are available in the following public GitHub repository: https://github.com/ShawanChowdhury/SocialMedia_RangeChange_TC.

Acknowledgments

We are grateful to the numerous volunteers for data collection and sharing their records on GBIF and social media. We collated the species distribution data from GBIF, Flickr and Facebook. The GBIF data (https://doi.org/10.15468/dl.wq7282) are publicly available. We have provided the Flickr and Facebook data in the supplementary data (Supplementary Data S1).

SC and AB gratefully acknowledge the support of the German Centre for Integrative Biodiversity Research (iDiv) and the sMon project funded by the German Research Foundation (through grants no. DFG-FZT 118 and no. 202548816).

Reference

Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., & Hegewisch, K. C. (2018). TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific Data*, *5*(1), 1-12.

Auffret, A. G., & Svenning, J. C. (2022). Climate warming has compounded plant responses to habitat conversion in northern Europe. *Nature Communications*, *13*(1), 7818.

Bergman, J. N., Buxton, R. T., Lin, H. Y., Lenda, M., Attinello, K., Hajdasz, A. C., Rivest, S. A., Tran Nguyen, T., & Bennett, J. R. (2022). Evaluating the benefits and risks of social media for wildlife conservation. *Facets*, *7*(1), 360–397.

Blackburn, T. M., Lockwood, J. L., & Cassey, P. (2016). The influence of numbers on invasion success. *Invasion Genetics: The Baker and Stebbins Legacy*, 25-39.

Braby, M. F., Bertelsmeier, C., Sanderson, C., & Thistleton, B. M. (2014). Spatial distribution and range expansion of the tawny Coster butterfly, Acraea terpsicore (Linnaeus, 1758)(Lepidoptera: Nymphalidae), in South-East Asia and Australia. *Insect conservation and diversity*, *7*(2), 132-143.

Broennimann O, Di Cola V, Guisan A (2023). ecospat: Spatial Ecology Miscellaneous Methods. R package version 4.0.0 <https://CRAN.R-project.org/package=ecospat>.

Broennimann, O., Fitzpatrick, M. C., Pearman, P. B., Petitpierre, B., Pellissier, L., Yoccoz, N. G., ... & Guisan, A. (2012). Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and Biogeography*, *21*(4), 481-497.

Butchart, S. H., Walpole, M., Collen, B., Van Strien, A., Scharlemann, J. P., Almond, R. E., ... & Watson, R. (2010). Global biodiversity: indicators of recent declines. *Science*, *328*(5982), 1164-1168.

Caley, P., & Cassey, P. (2023). Do we need to mine social media data to detect exotic vertebrate-pest introductions?. *Wildlife Research*, *50*(11), 869-875.

Capinha, C., Ceia-Hasse, A., de-Miguel, S., Vila-Viçosa, C., Porto, M., Jarić, I., ... & Pereira, H. M. (2024). Using citizen science data for predicting the timing of ecological phenomena across regions. *BioScience*, biae041.

Cardoso, A. S., Malta-Pinto, E., Tabik, S., August, T., Roy, H. E., Correia, R., ... & Vaz, A. S. (2024). Can citizen science and social media images support the detection of new invasion sites? A deep learning test case with *Cortaderia selloana*. *Ecological Informatics*, *81*, 102602.

Castro, A., Pinto, J., Reino, L., Pipek, P., & Capinha, C. (2024). Large language models overcome the challenges of unstructured text data in ecology. *Ecological Informatics*, *82*, 102742.

Chan, W. P., Lenoir, J., Mai, G. S., Kuo, H. C., Chen, I. C., & Shen, S. F. (2024). Climate velocities and species tracking in global mountain regions. *Nature*, *629*(8010), 114-120.

Chandler, M., See, L., Copas, K., Bonde, A. M., López, B. C., Danielsen, F., Legind, J. K., Masinde, S., Miller-Rushing, A. J., Newman, G., Rosemartin, A., & Turak, E. (2017). Contribution of citizen science towards international biodiversity monitoring. *Biological Conservation*, *213*, 280–294.

Chen, I. C., Hill, J. K., Ohlemüller, R., Roy, D. B., & Thomas, C. D. (2011). Rapid range shifts of species associated with high levels of climate warming. *Science*, *333*(6045), 1024-1026.

Chowdhury, S., Braby, M. F., Fuller, R. A., & Zalucki, M. P. (2021a). Coasting along to a wider range: niche conservatism in the recent range expansion of the Tawny Coster, Acraea terpsicore (Lepidoptera: Nymphalidae). *Diversity and Distributions*, *27*(3), 402-415.

Chowdhury, S., Fuller, R. A., Dingle, H., Chapman, J. W., & Zalucki, M. P. (2021b). Migration in butterflies: a global overview. *Biological Reviews*, *96*(4), 1462-1483.

Chowdhury, S. (2023). Threatened species could be more vulnerable to climate change in tropical countries. *Science of the Total Environment*, *858*, 159989.

Chowdhury, S., Aich, U., Rokonuzzaman, M., Alam, S., Das, P., Siddika, A., ... & Callaghan, C. T. (2023a). Increasing biodiversity knowledge through social media: A case study from tropical Bangladesh. *BioScience*, *73*(6), 453-459.

Chowdhury, S., Fuller, R. A., Rokonuzzaman, M., Alam, S., Das, P., Siddika, A., ... & Hanson, J. O. (2023b). Insights from citizen science reveal priority areas for conserving biodiversity in Bangladesh. *One Earth*, *6*(10), 1315-1325.

Chowdhury, S., Fuller, R. A., Ahmed, S., Alam, S., Callaghan, C. T., Das, P., ... & Bonn, A. (2024a). Using social media records to inform conservation planning. *Conservation Biology*, *38*(1), e14161.

Chowdhury, S., Ahmed, S., Alam, S., Callaghan, C. T., Das, P., Di Marco, M., ... & Bonn, A. (2024b). A protocol for harvesting biodiversity data from Facebook. *Conservation Biology*, e14257.

Convention on Biological Diversity (CBD). (2022). Kunming–Montreal Global biodiversity framework. Draft decision submitted by the President CBD/COP/15/L.25, 18 December 2022. https://www.cbd.int/conferences/2021-2022/cop-15/documents.

Correia, R. A., Ladle, R., Jarić, I., Malhado, A. C., Mittermeier, J. C., Roll, U., Soriano-Redondo, A., Veríssimo, D., Fink, C., Hausmann, A., Guedes-Santos, J., Vardi, R., & Di Minin, E. (2021). Digital data sources and methods for conservation culturomics. *Conservation Biology*, *35*(2), 398–411.

Dawson, N. M., Coolsaet, B., Bhardwaj, A., Brown, D., Lliso, B., Loos, J., ... & Worsdell, T. (2024). Reviewing the science on 50 years of conservation: Knowledge production biases and lessons for practice. *Ambio*, 1-19. https://doi.org/10.1007/s13280-024-02049-w.

Díaz, S., Settele, J., Brondízio, E. S., Ngo, H. T., Agard, J., Arneth, A., ... & Zayas, C. N. (2019). Pervasive human-driven decline of life on Earth points to the need for transformative change. *Science*, *366*(6471), eaax3100.

Di Marco, M., Chapman, S., Althor, G., Kearney, S., Besancon, C., Butt, N., ... & Watson, J. E. (2017). Changing trends and persisting biases in three decades of conservation science. *Global Ecology and Conservation*, *10*, 32-42.

Di Marco, M., Pacifici, M., Maiorano, L., & Rondinini, C. (2021). Drivers of change in the realised climatic niche of terrestrial mammals. *Ecography*, *44*(8), 1180-1190.

Di Minin, E., Fink, C., Hausmann, A., Kremer, J., & Kulkarni, R. (2021). How to address data privacy concerns when using social media data in conservation science. *Conservation Biology*, *35*(2), 437–446.

Di Minin, E., Selier, J., Louis, M., & Bradshaw, C. J. (2022). Dismantling the poachernomics of the illegal wildlife trade. *Biological Conservation*, *265*, Article 109418.

Di Minin, E., Tenkanen, H., & Toivonen, T. (2015). Prospects and challenges for social media data in conservation science. *Frontiers in Environmental Science*, *3*, Article 63.

Dirzo, R., Young, H. S., Galetti, M., Ceballos, G., Isaac, N. J., & Collen, B. (2014). Defaunation in the Anthropocene. *Science*, *345*(6195), 401-406.

Elith, J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution*, *1*(4), 330-342.

Feeley, K. J., Stroud, J. T., & Perez, T. M. (2017). Most 'global' reviews of species' responses to climate change are not truly global. *Diversity and Distributions*, *23*(3), 231-234.

Freeman, B. G., Scholer, M. N., Ruiz-Gutierrez, V., & Fitzpatrick, J. W. (2018). Climate change causes upslope shifts and mountaintop extirpations in a tropical bird community. *Proceedings of the National Academy of Sciences*, *115*(47), 11982-11987.

GBIF.org (7 July 2024) GBIF Occurrence Download https://doi.org/10.15468/dl.wq7282.

Guisan, A., Petitpierre, B., Broennimann, O., Daehler, C., & Kueffer, C. (2014). Unifying niche shift studies: insights from biological invasions. *Trends in Ecology & Evolution*, *29*(5), 260-269.

Hausmann, A., Toivonen, T., Slotow, R., Tenkanen, H., Moilanen, A., Heikinheimo, V., & Di Minin, E. (2018). Social media data can be used to understand tourists' preferences for nature-based experiences in protected areas. *Conservation Letters*, *11*(1), e12343.

Heberling, J. M., Miller, J. T., Noesgaard, D., Weingart, S. B., & Schigel, D. (2021). Data integration enables global biodiversity synthesis. *Proceedings of the National Academy of Sciences*, *118*(6), e2018093118.

Hijmans R. (2022). geosphere: Spherical Trigonometry. R package version 1.5-18, <https://CRAN.R-project.org/package=geosphere>.

Hijmans R. (2023). _raster: Geographic Data Analysis and Modeling_. R package version 3.6-26, <https://CRAN.R-project.org/package=raster>.

Hodgson, J. A., Thomas, C. D., Dytham, C., Travis, J. M., & Cornell, S. J. (2012). The speed of range shifts in fragmented landscapes.

Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven shortfalls that beset large-scale knowledge of biodiversity. Annual Review of Ecology, Evolution, and Systematics, 46(1), 523-549.

Hughes, A. C., Orr, M. C., Ma, K., Costello, M. J., Waller, J., Provoost, P., ... & Qiao, H. (2021). Sampling biases shape our view of the natural world. *Ecography*, *44*(9), 1259-1269.

Iseli, E., Chisholm, C., Lenoir, J., Haider, S., Seipel, T., Barros, A., ... & Alexander, J. M. (2023). Rapid upwards spread of non-native plants in mountains across continents. *Nature Ecology & Evolution*, *7*(3), 405-413.

Jarić, I., Correia, R. A., Brook, B. W., Buettel, J. C., Courchamp, F., Di Minin, E., ... & Roll, U. (2020). iEcology: harnessing large online resources to generate ecological insights. *Trends in Ecology & Evolution*, *35*(7), 630-639.

Jaureguiberry, P., Titeux, N., Wiemers, M., Bowler, D. E., Coscieme, L., Golden, A. S., ... & Purvis, A. (2022). The direct drivers of recent global anthropogenic biodiversity loss. *Science Advances*, *8*(45), eabm9982.

Joppa, L. N., O'Connor, B., Visconti, P., Smith, C., Geldmann, J., Hoffmann, M., ... & Burgess, N. D. (2016). Filling in biodiversity threat gaps. *Science*, *352*(6284), 416-418.

Kass, J. M., Muscarella, R., Galante, P. J., Bohl, C. L., Pinilla-Buitrago, G. E., Boria, R. A., ... & Anderson, R. P. (2021). ENMeval 2.0: Redesigned for customizable and reproducible modeling of species' niches and distributions. *Methods in Ecology and Evolution*, *12*(9), 1602-1608.

Kuussaari, M., Saccheri, I., Camara, M., & Hanski, I. (1998). Allee effect and population dynamics in the Glanville fritillary butterfly. *Oikos*, 384-392.

Lawlor, J. A., Comte, L., Grenouillet, G., Lenoir, J., Baecher, J. A., Bandara, R. M. W. J., ... & Sunday, J. (2024). Mechanisms, detection and impacts of species redistributions under climate change. N*ature Reviews Earth & Environment*, 1-18. https://www.nature.com/articles/s43017-024-00527-z.

Lenoir, J., & Svenning, J. C. (2015). Climate-related range shifts–a global multidimensional synthesis and new research directions. *Ecography*, *38*(1), 15-28.

Lenoir, J., Bertrand, R., Comte, L., Bourgeaud, L., Hattab, T., Murienne, J., & Grenouillet, G. (2020). Species better track climate warming in the oceans than on land. *Nature Ecology & Evolution*, *4*(8), 1044-1059.

Marcenò, C., Padullés Cubino, J., Chytrý, M., Genduso, E., Salemi, D., La Rosa, A., ... & Guarino, R. (2021). Facebook groups as citizen science tools for plant species monitoring. *Journal of Applied Ecology*, *58*(10), 2018-2028.

Maxwell, S. L., Fuller, R. A., Brooks, T. M., & Watson, J. E. (2016). Biodiversity: The ravages of guns, nets and bulldozers. *Nature*, *536*(7615), 143-145.

McCain, C. M., & Garfinkel, C. F. (2021). Climate change and elevational range shifts in insects. *Current Opinion in Insect Science*, *47*, 111-118.

Mesaglio, T., & Callaghan, C. T. (2021). An overview of the history, current contributions and future outlook of iNaturalist in Australia. *Wildlife Research*, *48*(4), 289-303.

Mota, L. L., Boddington, S. J., Brown Jr, K. S., Callaghan, C. J., Carter, G., Carter, W., ... & Freitas, A. V. (2022). The butterflies of Cristalino Lodge, in the Brazilian southern Amazonia: An updated species list with a significant contribution from citizen science. *Biota Neotropica*, *22*(3), e20221367.

Nowakowski, A. J., Watling, J. I., Whitfield, S. M., Todd, B. D., Kurz, D. J., & Donnelly, M. A. (2017). Tropical amphibians in shifting thermal landscapes under land-use and climate change. *Conservation Biology*, *31*(1), 96-105.

Parker, E. J., Weiskopf, S. R., Oliver, R. Y., Rubenstein, M. A., & Jetz, W. (2024). Insufficient and biased representation of species geographic responses to climate change. *Global Change Biology*, *30*, e17408.

Pernat, N., Canavan, S., Golivets, M., Hillaert, J., Itescu, Y., Jarić, I., Mann, H. M. R., Pipek, P., Preda, C., Richardson, D. M., Teixeira, H., Vaz, A. S., & Groom, Q. (2024). Overcoming biodiversity blindness: Secondary data in primary citizen science observations. *Ecological Solutions and Evidence*, *5*, e12295.

Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E., & Blair, M. E. (2017). Opening the black box: An open-source release of Maxent. *Ecography*, *40*(7), 887-893.

Pimm, S. L., Jenkins, C. N., Abell, R., Brooks, T. M., Gittleman, J. L., Joppa, L. N., ... & Sexton, J. O. (2014). The biodiversity of species and their rates of extinction, distribution, and protection. *Science*, *344*(6187), 1246752.

Pocock, M. J., Chandler, M., Bonney, R., Thornhill, I., Albin, A., August, T., ... & Danielsen, F. (2018). A vision for global biodiversity monitoring with citizen science. In *Advances in Ecological Research* (Vol. 59, pp. 169-223). Academic Press.

Poloczanska, E. S., Burrows, M. T., Brown, C. J., García Molinos, J., Halpern, B. S., Hoegh-Guldberg, O., ... & Sydeman, W. J. (2016). Responses of marine organisms to climate change across oceans. *Frontiers in Marine Science*, *3*, 62.

Pounds, A. J., Bustamante, M. R., Coloma, L. A., Consuegra, J. A., Fogden, M. P., Foster, P. N., ... & Young, B. E. (2006). Widespread amphibian extinctions from epidemic disease driven by global warming. *Nature*, *439*(7073), 161-167.

R Core Team (2024). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Rushing, C. S., Royle, J. A., Ziolkowski Jr, D. J., & Pardieck, K. L. (2020). Migratory behavior and winter geography drive differential range shifts of eastern birds in response to recent climate change. *Proceedings of the National Academy of Sciences*, *117*(23), 12897-12903.

Sanderson, C., Braby, M. F., Thistleton, B. M., & Neal, M. (2012). First record of the Tawny Coster butterfly Acraea terpsicore (Linnaeus, 1758) from Australia. *Myrmecia. News Bulletin of the Australian Entomological Society*, *48*, 39–41.

Santos, M. J., Smith, A. B., Thorne, J. H., & Moritz, C. (2017). The relative influence of change in habitat and climate on elevation range limits in small mammals in Yosemite National Park, California, USA. *Climate Change Responses*, *4*, 1-12.

Sbragaglia, V., Espasandín, L., Jarić, I., Vardi, R., Ramírez, F. and Coll, M. (2024). Tracking ongoing transboundary marine distributional range shifts in the digital era. *Marine Ecology Progress Series*, *728*, 103-114.

Soriano-Redondo, A., Correia, R. A., Barve, V., Brooks, T. M., Butchart, S. H., Jarić, I., ... & Di Minin, E. (2024). Harnessing online digital data in biodiversity monitoring. *PLoS Biology*, *22*(2), e3002497.

Sheard, J.K., Adriaens, T., Bowler, D.E., Büermann, A., Callaghan, C.T., Camprasse, E.C.M., Chowdhury, S., Engel, T., Finch, E.A., von Gönner, J., Hsing, P.-Y., Mikula, P., Oh, R.Y.R., Peters, B., Phartyal, S.S., Pocock, M.J.O., Wäldchen, J. & Bonn, A. (2024) Emerging technologies in citizen science and potential for insect monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *379*, 20230106.

Toivonen, T., Heikinheimo, V., Fink, C., Hausmann, A., Hiippala, T., Järv, O., ... & Di Minin, E. (2019). Social media data for conservation science: A methodological overview. *Biological Conservation*, *233*, 298-315.

van Klink, R., August, T., Bas, Y., Bodesheim, P., Bonn, A., Fossøy, F., Høye, T.T., Jongejans, E., Menz, M.H.M., Miraldo, A., Roslin, T., Roy, H.E., Ruczyński, I., Schigel, D., Schäffler, L., Sheard, J.K., Svenningsen, C., Tschan, G.F., Wäldchen, J., Zizka, V.M.A., Åström, J. & Bowler, D.E. (2022) Emerging technologies revolutionise insect ecology and monitoring. *Trends in Ecology & Evolution*, *37*, 872-885.

Warren, D. L., Glor, R. E., & Turelli, M. (2008). Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution*, *62*(11), 2868-2883.

Whitford, A. M., Shipley, B. R., & McGuire, J. L. (2024). The influence of the number and distribution of background points in presence-background species distribution models. *Ecological Modelling*, *488*, 110604.

Wiens, J. J., Ackerly, D. D., Allen, A. P., Anacker, B. L., Buckley, L. B., Cornell, H. V., ... & Stephens, P. R. (2010). Niche conservatism as an emerging principle in ecology and conservation biology. *Ecology Letters*, *13*(10), 1310-1324.

Wilson, E. O. (1989). Threats to biodiversity. *Scientific American*, *261*(3), 108-117.

Yackulic, C. B., Sanderson, E. W., & Uriarte, M. (2011). Anthropogenic and environmental drivers of modern range loss in large mammals. *Proceedings of the National Academy of Sciences*, *108*(10), 4024-4029.

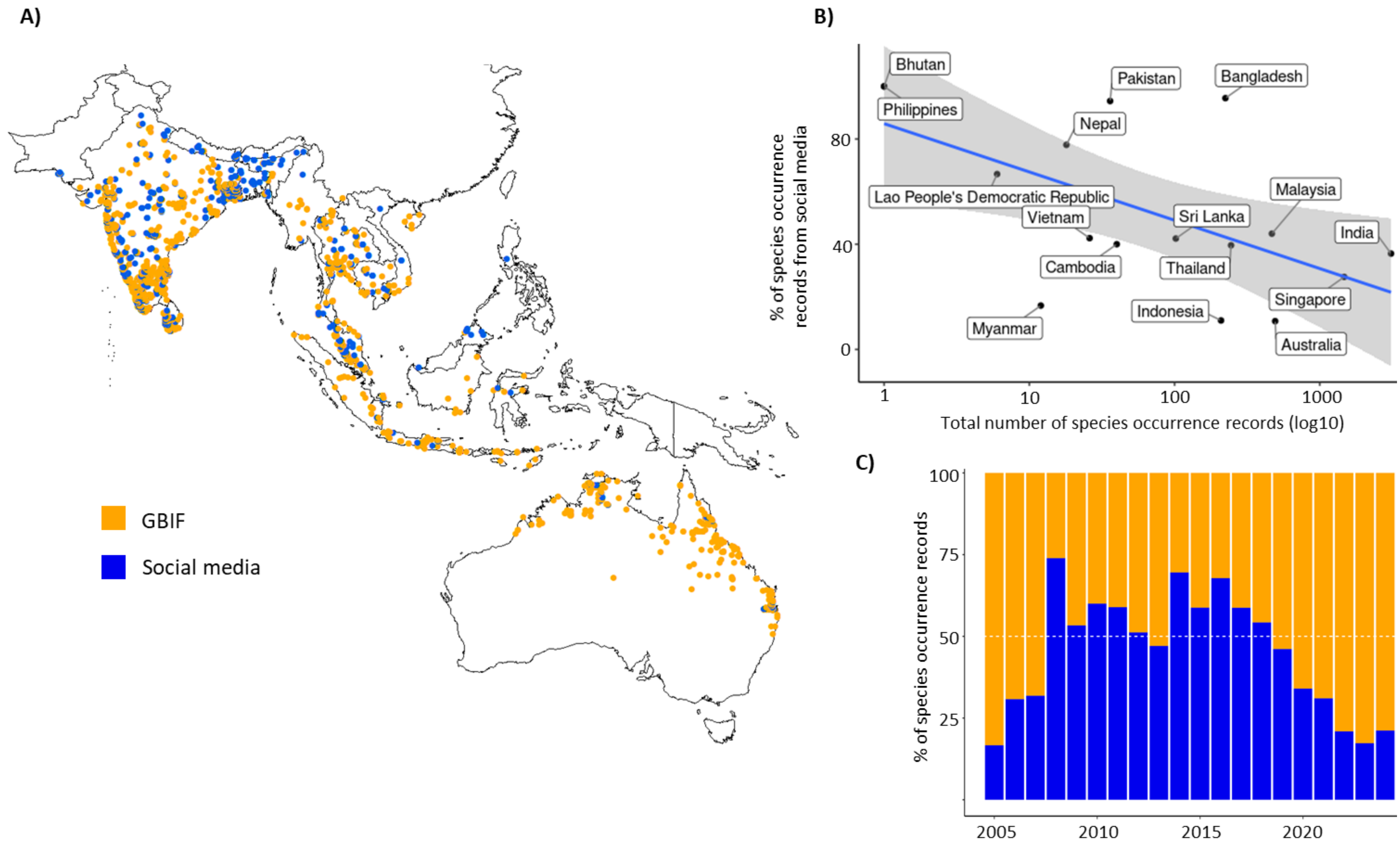
List of Figures

Figure 1. Distribution of occurrence records by source (GBIF and social media, where social media contains records from both Facebook and Flickr) with (A) the entire known distribution of tawny coster (*Acraea terpsicore*) (each point represents one occurrence record); (B) the association between the total number of occurrence records and the percentages of occurrences records obtained from social media data; and (C) temporal trends in occurrence records by data source (GBIF vs. social media).

Figure 2. Tracking range shift dynamics of the tawny coster butterfly (*Acraea terpsicore*) during 2005-2024. (A-E) Maps of predicted habitat suitability in five-year intervals, highlighting areas where model predictions altered after adding the social media data (i.e., contribution from social media data in model predictions is highlighted in purple); (F) the differences in the surface area obtained from binarized model predictions; (G) latitudinal and elevational distribution of the suitable areas (for all year intervals); and (H) the differences in estimated range expansion.

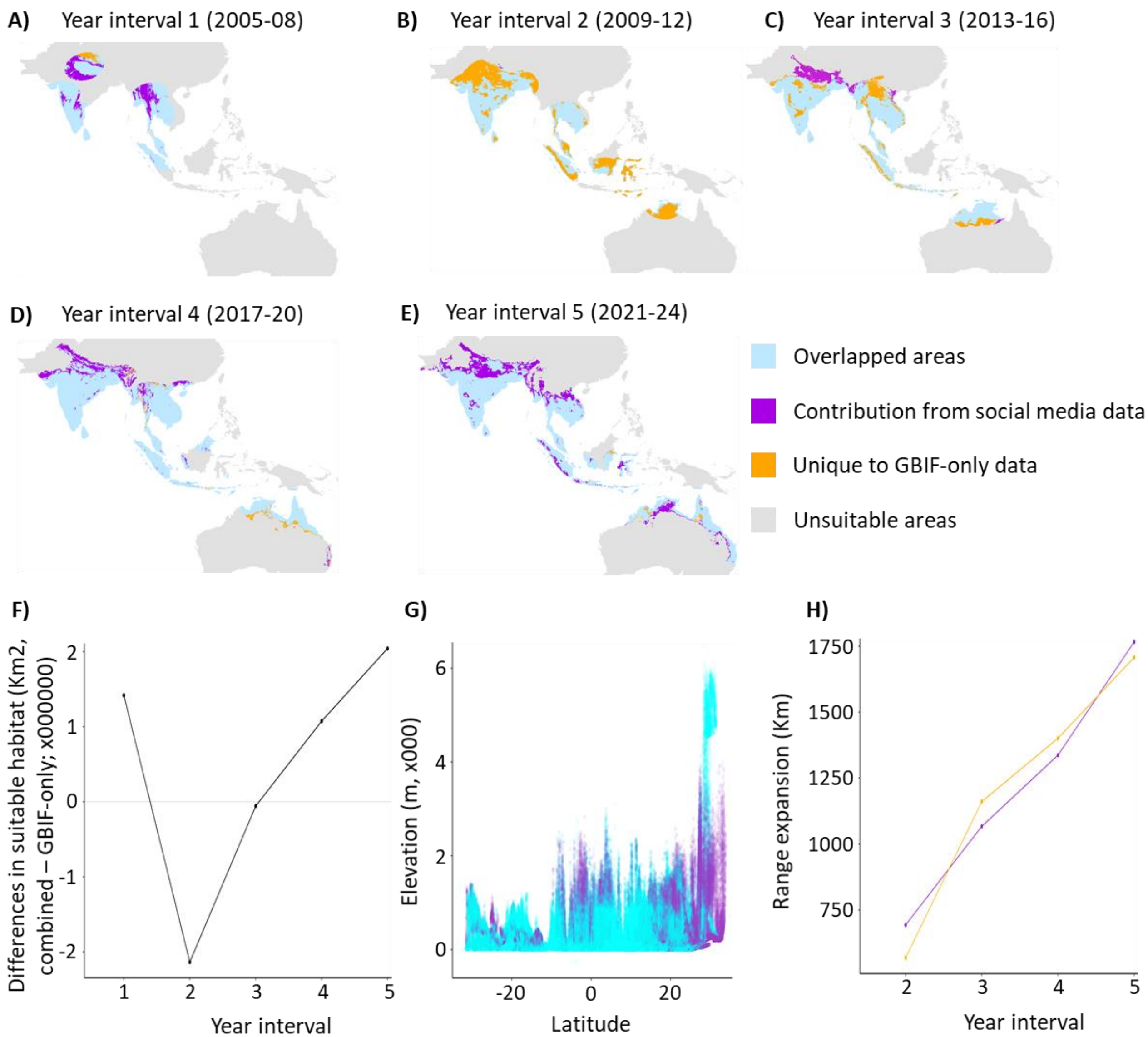
Figure 3. The differences in identified niche space with adding social media data in different year intervals. The inset figures (correlation plot) show the importance and direction of impact of each predictor variable. Here, def = climate water deficit, ppt = precipitation (monthly total), soil = soil moisture, tmax = maximum temperature, ws = wind speed, PDSI = Palmer Drought Severity Index, and elev = elevation.

Figure 1.



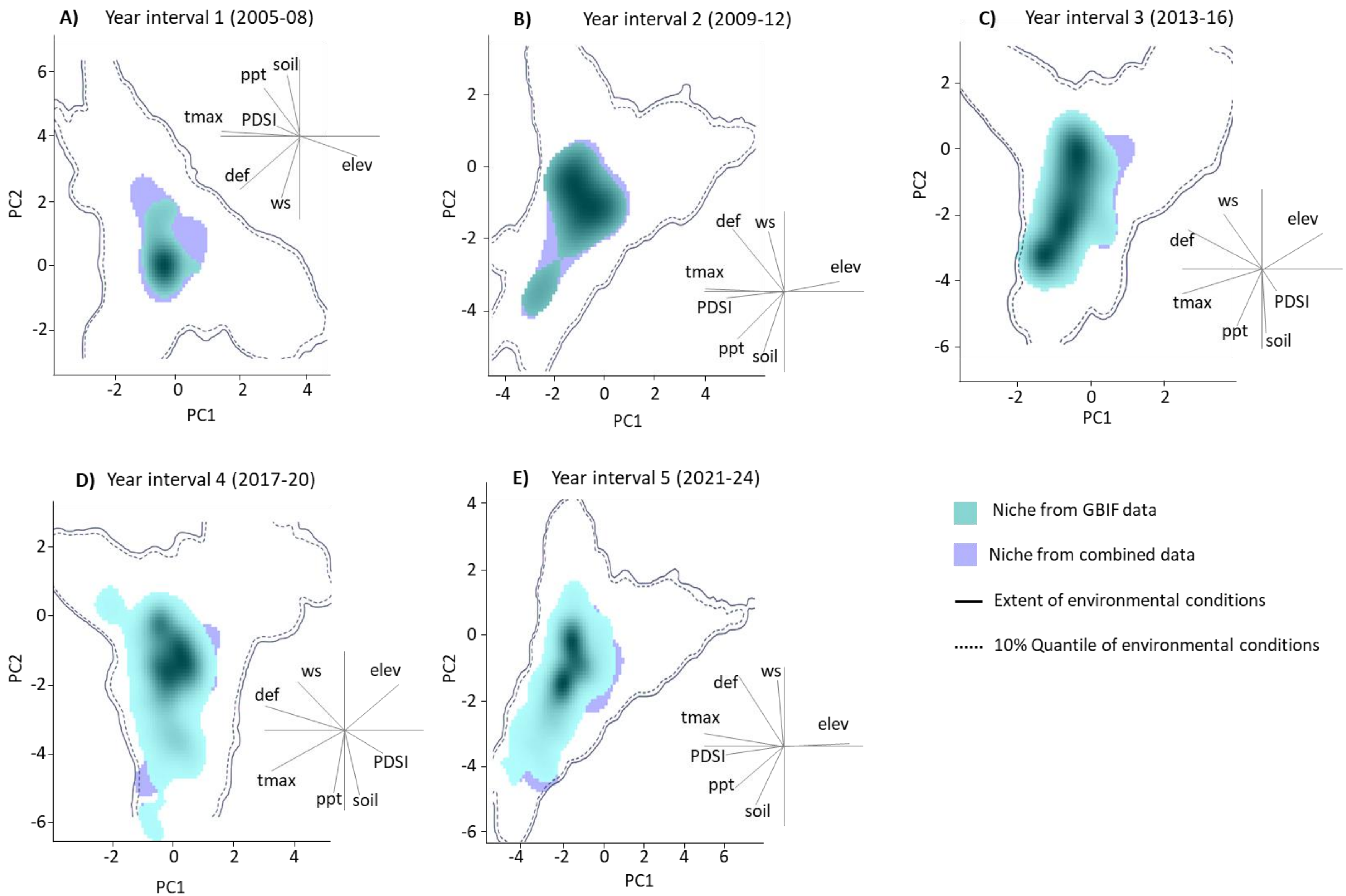
A)
GBIF
Social media
B)
% of species occurrence records from social media
Total number of species occurrence records (log10)
0
40
80
1
10
100
1000
Bhutan
Philippines
Pakistan
Bangladesh
Nepal
Lao People's Democratic Republic
Vietnam
Sri Lanka
Malaysia
India
Cambodia
Thailand
Singapore
Myanmar
Indonesia
Australia
C)
% of species occurrence records
25
50
75
100
2005
2010
2015
2020

Figure 2.



A) Year interval 1 (2005-08)
B) Year interval 2 (2009-12)
C) Year interval 3 (2013-16)
D) Year interval 4 (2017-20)
E) Year interval 5 (2021-24)
Overlapped areas
Contribution from social media data
Unique to GBIF-only data
Unsuitable areas
F)
Differences in suitable habitat (Km2, combined – GBIF-only; x000000)
Year interval
G)
Elevation (m, x000)
Latitude
H)
Range expansion (Km)
Year interval

633 **Figure 3.**



634

Supplementary Table S1: Details of the Facebook group, showing the keywords used, country information, group names, and URL. The country information NA means that the group is not country-specific.

Keywords	Country	Group Name	Group URL
Butterfly Bangladesh	Bangladesh	Butterfly Bangladesh	https://www.facebook.com/groups/butterflybangladesh/
Butterfly India	India	ButterflyIndia	https://www.facebook.com/groups/ButterflyIndia/
Butterfly India	NA	World of Butterflies	https://www.facebook.com/groups/208726699864915/
Butterfly India	India	Indian Butterflies	https://www.facebook.com/groups/indianbutterflies/
Butterflies India	India	Moths and Butterflies of Northeast India	https://www.facebook.com/groups/654169375175548/
Butterflies India	India	Butterflies Of West Bengal	https://www.facebook.com/groups/224547761077777/
Butterflies India	India	Butterflies of North-Eastern India	https://www.facebook.com/groups/butterfliesofnortheastindia/
Butterflies India	India	Insects and butterflies of India	https://www.facebook.com/groups/1609109409365987/
Butterflies India	India	Moths and Butterflies of Northeast India	https://www.facebook.com/groups/654169375175548/?hoisted_section_header_type=recently_seen&multi_permalinks=1496482467610897
Butterflies India	India	Butterflies From India	https://www.facebook.com/groups/butterfliesfromindia/
Butterflies India	India	Butterflies of India	https://www.facebook.com/groups/246930848714418/?hoisted_section_header_type=recently_seen&multi_permalinks=8239610469446376

Butterfly Australia	Australia	Butterflies, Moths and other Invertebrates of Australia	https://www.facebook.com/groups/170745013686340/
Butterfly Australia	Australia	Australian Butterflies	https://www.facebook.com/groups/1642134862743621/
Butterfly Australia	Australia	Australian butterflies and moths	https://www.facebook.com/groups/799465170167144/
Butterfly India	India	InsectIndia	https://www.facebook.com/groups/InsectIndia/
Butterfly India	India	Ask IDs of Indian Butterflies	https://www.facebook.com/groups/275650439625884/
Butterfly India	NA	Australasian Butterflies and Moths	https://www.facebook.com/groups/305768640902911/
Butterfly Pakistan	Pakistan	Butterflies of Pakistan	https://www.facebook.com/groups/131718433700946/
Butterfly Bhutan	Bhutan	Butterflies Society of Bhutan	https://www.facebook.com/groups/2287652334879957/
Butterfly Bhutan	Bhutan	Butterfly and Moths of Bhutan	https://www.facebook.com/groups/bhutanmoths/
Butterfly Nepal	Nepal	Butterfly Conservation Nepal.NP	https://www.facebook.com/groups/hinalayanbirder/
Butterfly Nepal	Nepal	Butterflies Diversity of Nepal	https://www.facebook.com/groups/1979680945491783/
Butterfly Nepal	Nepal	Butterflies of Nepal	https://www.facebook.com/groups/butterfliesnepal/
Butterfly China	NA	Butterflying Around the World 寰宇蝶影	https://www.facebook.com/groups/649414648579768/
Butterfly Australia	Australia	South-east Queensland Butterfly Watching	https://www.facebook.com/groups/seqbutterflies/

Butterflies Bangladesh	Bangladesh	Butterflies of Bangladesh	https://www.facebook.com/groups/129902820822634/
Butterflies Bangladesh	Bangladesh	ButterflyBengal	https://www.facebook.com/groups/333424654047804/
Butterfly SriLanka	SriLanka	Butterfly Conservation & Research Group of Sri Lanka	https://www.facebook.com/groups/bcrgsl/
Butterflies Myanmar	Myanmar	Butterflies and Moths of Myanmar	https://www.facebook.com/groups/624495890937290/
Butterflies Thailand	Thailand	Butterflies of Thailand	https://www.facebook.com/groups/365316596902175/?hoisted_section_header_type=recently_seen&multi_permalinks=2977081182392357
Butterfly Cambodia	Cambodia	Natural Cambodia	https://www.facebook.com/groups/naturalcambodia/
Butterfly Singapore	Singapore	Butterfly & Macro Singapore	https://www.facebook.com/groups/1885942558332507/
Butterfly Singapore	Singapore	ButterflyCircle (Butterflies of Singapore)	https://www.facebook.com/groups/240038746511844/
Butterfly Singapore	Singapore	Singapore (SG) Insect ID and Records	https://www.facebook.com/groups/sginsectid/
Butterflies Singapore	Singapore	Butterflies of Singapore and Malaysia	https://www.facebook.com/groups/255047171183480/?hoisted_section_header_type=recently_seen&multi_permalinks=7776253999062722
Butterfly Taiwan	Taiwan	台灣超微距暨昆蟲攝影交流會 Taiwan Super Macro	https://www.facebook.com/groups/1416552225029363/?hoisted_section_header_type=recently_seen&multi_permalinks=2556234093357155

Butterfly Indonesia	Indonesia	Indonesia Photography Family	https://www.facebook.com/groups/950185375659760/?hoisted_section_header_type=recently_seen&multi_permalinks=1346355336042760
Butterflies Indonesia	Indonesia	Butterflies of Borneo	https://www.facebook.com/groups/1637283956484389/
Butterflies Philippines	Philippines	Philippine Lepidoptera	https://www.facebook.com/groups/488909304537513/
Butterflies Philippines	Philippines	PARUPAROZZIS: Butterfly Watchers Philippines	https://www.facebook.com/groups/paruparozzi/
Butterflies Papua New Guinea	Papua New Guinea	Butterflies of Papua	https://www.facebook.com/groups/2185597161721882/

Supplementary Figure S2: The 10000 background maps that were randomly generated using all the occurrence records of the tawny coster butterfly.