



OPEN

DATA DESCRIPTOR

# Birds of a feather flock together: a dataset for *Clock* and *Adcyap1* genes from migration genetics studies

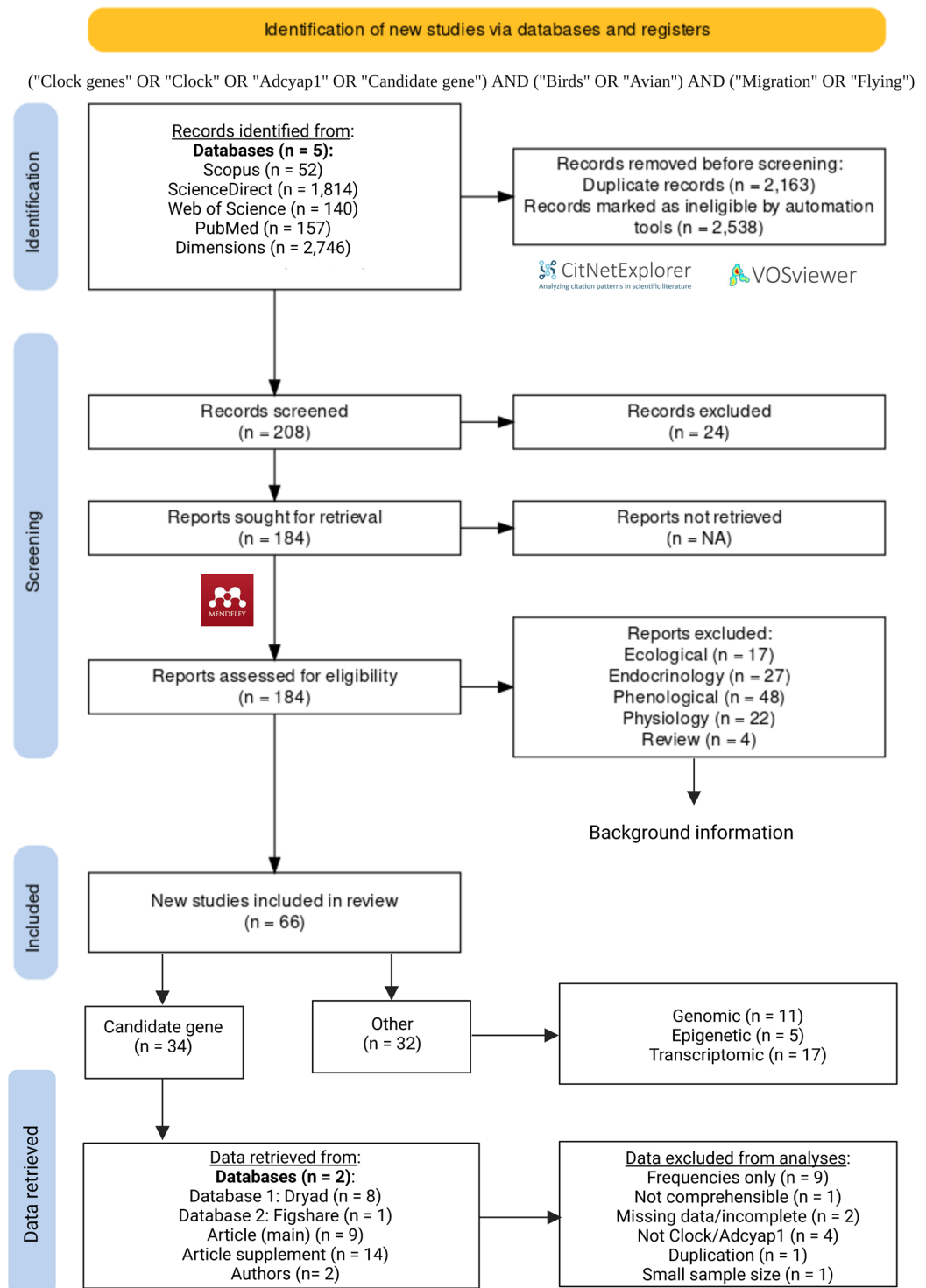
Louis-Stéphane Le Clercq<sup>1,2</sup>✉, Gaia Bazzi<sup>3</sup>, Joan Ferrer Obiol<sup>4,5,6</sup>, Jacopo G. Cecere<sup>3</sup>, Luca Gianfranceschi<sup>7</sup>, J. Paul Grobler<sup>2</sup>, Antoinette Kotzé<sup>1,2</sup>, Marta Riutort León<sup>4,5</sup>, Jacob González-Solís<sup>5,8</sup>, Diego Rubolini<sup>6,9</sup>, Miriam Liedvogel<sup>10,11</sup> & Desiré Lee Dalton<sup>12</sup>

Birds in seasonal habitats rely on intricate strategies for optimal timing of migrations. This is governed by environmental cues, including photoperiod. Genetic factors affecting intrinsic timekeeping mechanisms, such as circadian clock genes, have been explored, yielding inconsistent findings with potential lineage-dependency. To clarify this evidence, a systematic review and phylogenetic reanalysis was done. This descriptor outlines the methodology for sourcing, screening, and processing relevant literature and data. PRISMA guidelines were followed, ultimately including 66 studies, with 34 focusing on candidate genes at the genotype-phenotype interface. Studies were clustered using bibliographic coupling and citation network analysis, alongside scientometric analyses by publication year and location. Data was retrieved for allele data from databases, article supplements, and direct author communications. The dataset, version 1.0.2, encompasses data from 52 species, with 46 species for the *Clock* gene and 43 for the *Adcyap1* gene. This dataset, featuring data from over 8000 birds, constitutes the most extensive cross-species collection for these candidate genes, used in studies investigating gene polymorphisms and seasonal bird migration.

## Background & Summary

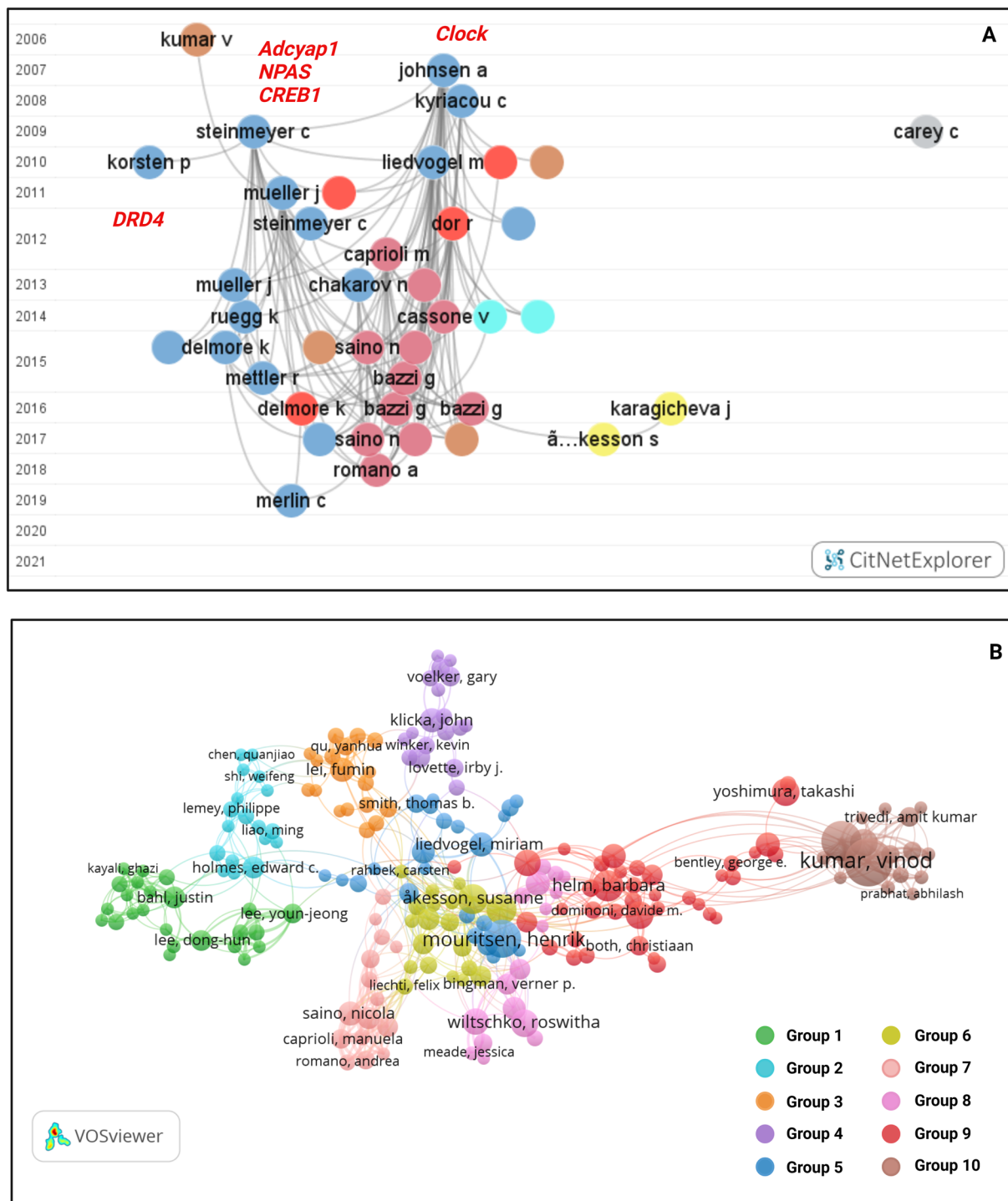
Birds occupy nearly every habitat and ecoregion on Earth, however, many of these habitats experience large seasonal shifts in key ecological attributes such as length of day<sup>1</sup>, temperature<sup>2</sup>, rainfall<sup>3,4</sup>, and associated food and nesting material availability<sup>5</sup>. This has necessitated the adaptive evolution of complex strategies to maximise survival through seasonal migrations between breeding and wintering ranges. Migrations are carefully timed events, scheduled in such a manner that birds can optimise hours of daylight<sup>6</sup>, nighttime visibility<sup>7,8</sup>, and time spent at stop-over sites<sup>9</sup> along their migration route to ensure timely arrivals for optimal habitat use. While most of the ecological attributes play some role in the timing of migration, one of the best studied attributes that serve as a trigger to initiate migration is the length of day or photoperiod. The photoperiod is primarily responsible for daily oscillations within the regulatory feedback loops of the circadian clock, which differentially expresses genes during light or dark phases to maintain sleep-wake cycles in most organisms<sup>10</sup>.

<sup>1</sup>South African National Biodiversity Institute, P.O. Box 754, Pretoria, 0001, South Africa. <sup>2</sup>Department of Genetics, University of the Free State, P.O. Box 339, Bloemfontein, 9300, South Africa. <sup>3</sup>Area Avifauna Migratrice, Istituto Superiore per la Protezione e la Ricerca Ambientale, via Ca' Fornacetta 9, I-40064, Ozzano Emilia, BO, Italy. <sup>4</sup>Departament de Genètica, Universitat de Barcelona, Gran Via de les Corts Catalanes, 585, 08007, Barcelona, Spain. <sup>5</sup>Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Gran Via de les Corts Catalanes, 585, 08007, Barcelona, Spain. <sup>6</sup>Dipartimento di Scienze e Politiche Ambientali, Università degli Studi di Milano, via Celoria 26, Milan, I-20133, Italy. <sup>7</sup>Dipartimento di Bioscienze, Università degli Studi di Milano, via Celoria 26, Milan, I-20133, Italy. <sup>8</sup>Departament de Biologia Evolutiva, Universitat de Barcelona, Gran Via de les Corts Catalanes, 585, 08007, Barcelona, Spain. <sup>9</sup>Istituto di Ricerca sulle Acque, IRSA-CNR, Via del Mulino 19, I-20861, Brugherio, (MB), Italy. <sup>10</sup>Max Planck Research Group Behavioural Genomics, Max Planck Institute for Evolutionary Biology, 24306, Plön, Germany. <sup>11</sup>Institute of Avian Research, An der Vogelwarte 21, 26386, Wilhelmshaven, Germany. <sup>12</sup>School of Health and Life Sciences, Teesside University, Middlesbrough, TS1 3BA, UK. ✉e-mail: [leclercq.l.s@gmail.com](mailto:leclercq.l.s@gmail.com)

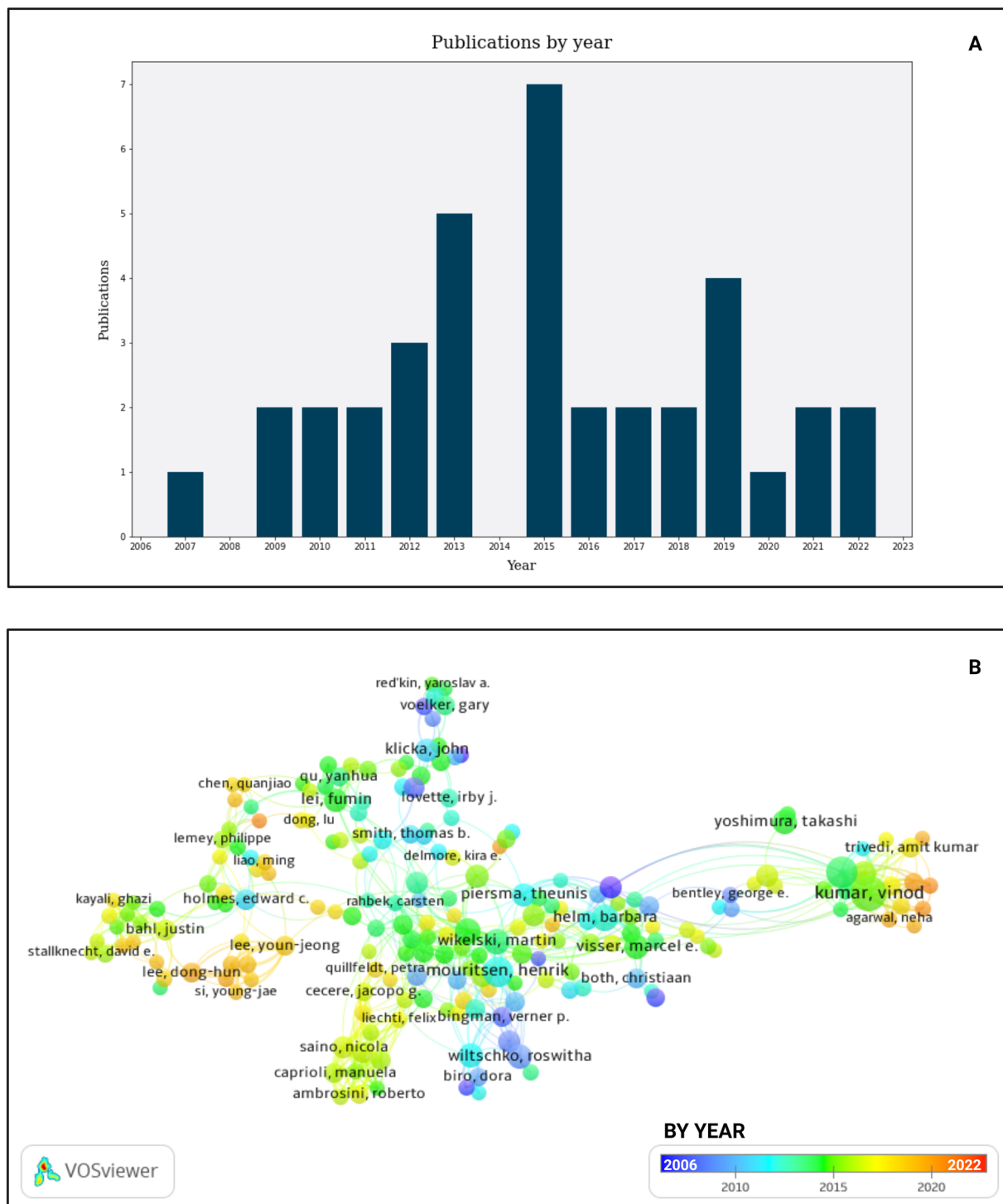


**Fig. 1** PRISMA statement for the systematic approach used to identify studies that measured clock gene polymorphisms in relation to annual synchronicity of live events such as breeding and migration in birds. Further details are also provided for the retrieval of allele data for individual studies from various sources as well as reasons for exclusion of studies. (image edited in BioRender.com).

One conundrum regarding migration in birds is how differential migration patterns are established and maintained within singular species, even in the absence of extrinsic environmental triggers. For example, several species within the order Coraciiformes have distinct populations that are either year-round residents, with minimal altitudinal movement, or long-distance migrants. This includes such species as the Lilac-breasted roller<sup>11</sup> (*Coracias caudatus*) and Woodland kingfisher<sup>12</sup> (*Halcyon senegalensis*), both having subspecies that are delineated by differential migration, as well as the European bee-eater<sup>13</sup> (*Merops apiaster*), which is considered

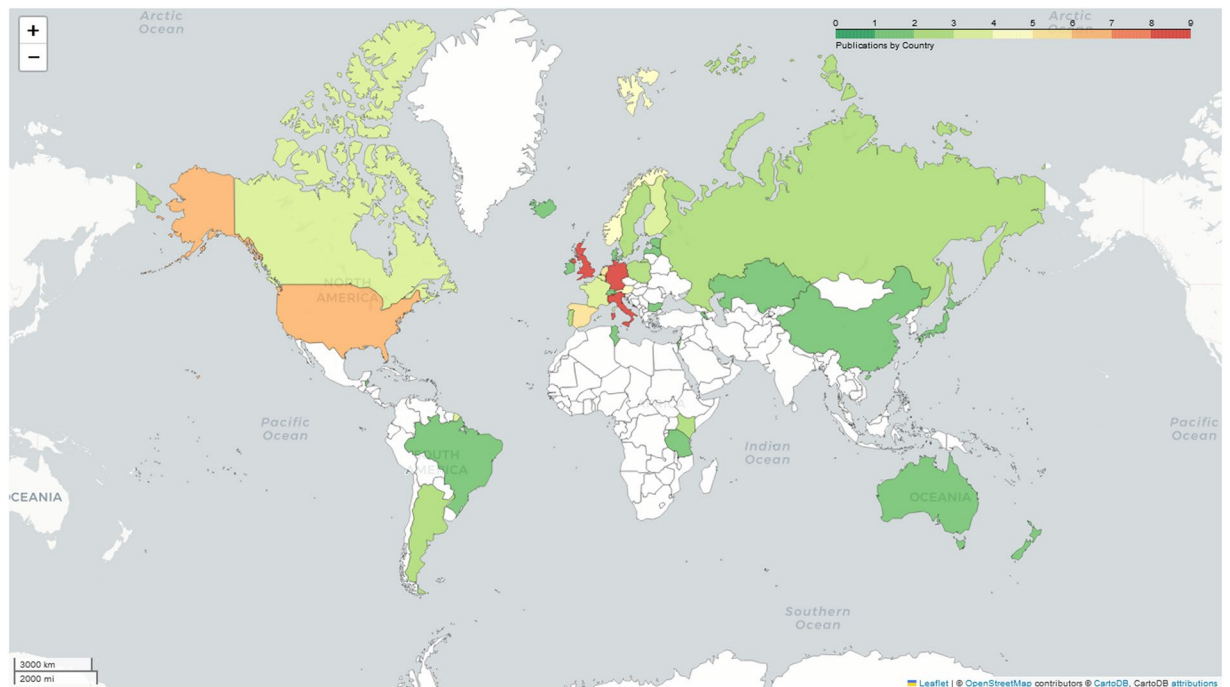


**Fig. 2** Visualised citation network for studies identified in literature searches. **(A)** Citation network of the Scopus and PubMed database in CitNetExplorer. Publications are organized by year (2006–2021) with the name and first initial of the first author indicating individual studies. The relationship between studies by virtue of co-citations in the reference lists are indicated by grey lines. Subgroup analyses identified several key groups, indicated by the colour code from VOSviewer. Key candidate genes are indicated in red italics and show studies that assayed polymorphisms in the *Clock*, *Adcyap1*, *CREB1*, *NPAS*, and *DRD4* genes. **(B)** Citation network for studies identified in literature searches of the Dimensions and ScienceDirect database in VOSviewer. First authors are labelled by surname and first name. Automated group analyses identified ten clusters of related studies of which the studies identified from Scopus formed part of five groups, indicated as groups 2, 5, 6, 7, 9, and 10. This network shows the larger field of migration studies including non-candidate gene studies such as transcriptomic studies (group 10). (image edited in BioRender.com).



**Fig. 3** Plots indicating the distribution for publications by year. **(A)** Histogram for publications by year indicating the first publications starting in 2007 up to more recent publications in 2022, with the largest number of publications between 2013–2015 and in 2019. **(B)** Density gradient display of studies in VOSviewer based on year of publication, indicated most studies were published between 2006 (blue) and 2022 (red) with a high number of publications emanating from 2013–2016 (green to orange). (image edited in BioRender.com).

monotypic but has a distinct resident population in Southern Africa. Understanding how differential migration is established and maintained between such species is key to assessing connectivity<sup>14</sup>, speciation at a subspecies level<sup>15</sup>, and potential population fitness<sup>16</sup>. This is particularly pertinent with regards to the plasticity or ability to switch between behaviours<sup>17,18</sup> should environmental conditions change considerably due to climate change<sup>19–21</sup> or anthropogenic activity<sup>22–25</sup>.



**Fig. 4** Geographic distribution of candidate gene studies included in the final review dataset ( $N = 34$ ) based on sampling locations. Related migration studies ( $N = 32$ ), such as transcriptomic or epigenetic studies, were excluded. The density gradient plots the number of studies per country ranging from one study (green) to more than eight studies (red); countries in white are data deficient. The overall plot indicates that most studies emanated from sampling locations in Europe and North America, with only a small number of studies including sampling from parts of Africa and South America.

Several studies have explored the possible genetic components that affect intrinsic time keeping mechanisms and migration. Although variable methods have been used, including genomic<sup>26</sup>, epigenetic<sup>27</sup>, and transcriptomic approaches<sup>28</sup>, most studies sought to identify genes or gene regions that show variation in either the sequence itself or the gene expression that can be correlated to divergent migratory behaviour. The key, however, is identifying variation that is linked to processes that interface with annual life events. Thus, variation that is either connected to the endocrine or metabolic changes<sup>29</sup>, in preparation for migration and breeding, or intrinsic time-keeping mechanisms, such as the rhythmic expression of circadian genes; particularly those that interface with environmental changes that may serve as cues such as photoperiod, temperature, lunar cycles, and food availability<sup>30</sup>. This is needed to exclude variants that co-vary with migration phenotypes but are not actively involved in shaping them. It is therefore no surprise that many candidate gene studies have explored variation within the network of genes of the circadian clock. Several associated candidate genes have been suggested, with length polymorphisms within short repeats of the *Clock* and *Adcyap1* genes being the focus of many studies<sup>31–33</sup>.

To clarify the role of these genes in migratory phenotypes, a systematic review (Fig. 1) was conducted to identify, synthesise, and provide a reappraisal of the available evidence<sup>34</sup>. Structured searches of the literature with an optimised Boolean search string were done in five scientific databases. Search results were exported in formats compatible with citation network analysis software<sup>35</sup>. After duplicate entries were removed, citation network analyses were used for the automated screening of database results to identify the central literature on the topic. Publications identified from the citation network analyses were subjected to manual screening of the title, abstract, and key words to assess the potential eligibility for inclusion in the review. The final list of most eligible publications was sought for full text retrieval. A total of 66 studies were included in the final review of which 34 were candidate gene studies and 32 were other, migration-related, studies. These included latitude/longitude/spatial analyses, timing of migration, and timing of egg laying/breeding. Most of the studies using a candidate gene approach were used for data retrieval. For these studies, datasets were retrieved as either diploid allele data of individuals or allele frequencies. Data sources included the main text of articles, supplementary materials, databases such as Dryad (<https://datadryad.org/>) or Figshare (<https://figshare.com/>), data extraction, or data received directly from authors. Unpublished data for an additional 12 species were also included. The dataset included individual level allele data from 52 species of which data was available for 46 species for the *Clock* gene and 43 species for the *Adcyap1* gene. This dataset represents the largest collection of cross species allele data for two candidate genes used to test a putative association between clock gene polymorphisms and divergent migration in birds, which enables the testing for patterns of inheritance, evolutionary selection, relation to divergence times, and associations across a globally distributed dataset.

This data descriptor summarises both the methodology used to screen the literature as well as to compile the data concisely and presents the resulting data used in prior analyses in an easy-to-understand format. At present, none of the scientific databases that collect genetic variation data is suitable for the deposit of this specific type

Common name	Latin binomial	Study	Data	Type	Location	N
Barn swallow*	<i>Hirundo rustica</i>	51,62,63	51	C <sub>A</sub>	Switzerland, Italy	64
Bar-tailed godwit	<i>Limosa lapponica baueri</i>	64	64	C <sub>A</sub>	New Zealand	135
Blackpoll warbler*	<i>Setophaga striata</i>	65	43	C <sub>A</sub> , A <sub>A</sub>	USA	72
Blue tit*	<i>Cyanistes caeruleus</i>	31,32,66,67	31,42,45	C <sub>A</sub>	Europe	950
Collared flycatcher*	<i>Ficedula albicollis</i>	68	47	C <sub>A</sub> , A <sub>A</sub>	Czechia	406
Collared plover	<i>Charadrius collaris</i>	69	69	A <sub>A</sub>	Brazil	14
Common buzzard	<i>Buteo buteo</i>	70	48	A <sub>A</sub>	Germany	978
Common nightingale*	<i>Luscinia megarhynchos</i>	54,71	44, Authors	C <sub>A</sub> , A <sub>A</sub>	Italy	150
Common redstart*	<i>Phoenicurus phoenicurus</i>	54	Authors	C <sub>A</sub> , A <sub>A</sub>	Italy	43
Common whitethroat*	<i>Sylvia communis</i>	54	Authors	C <sub>A</sub> , A <sub>A</sub>	Italy	25
Dark-eyed junco*	<i>Junco hyemalis</i>	72	72	C <sub>A</sub> , A <sub>A</sub>	USA	36
Eastern subalpine warbler*	<i>Currucacantillans</i>	54	Authors	C <sub>A</sub> , A <sub>A</sub>	Italy	31
Eurasian blackbird*	<i>Turdus merula</i>	22	22	C <sub>A</sub> , A <sub>A</sub>	Europe, Tunisia	792
Eurasian blackcap*	<i>Sylvia atricapilla</i>	73,74	50,73	A <sub>A</sub>	Europe	936
Eurasian golden oriole*	<i>Oriolus oriolus</i>	54	Authors	C <sub>A</sub> , A <sub>A</sub>	Italy	30
Eurasian hoopoe*	<i>Upupa epops</i>	54	Authors	C <sub>A</sub> , A <sub>A</sub>	Italy	25
Eurasian reed warbler*	<i>Acrocephalus scirpaceus</i>	54	Authors	C <sub>A</sub> , A <sub>A</sub>	Italy	24
Eurasian wryneck*	<i>Jynx torquilla</i>	54	Authors	C <sub>A</sub> , A <sub>A</sub>	Italy	30
European bee-eater*	<i>Merops apiaster</i>	54	Authors	C <sub>A</sub> , A <sub>A</sub>	Italy	35
European nightjar*	<i>Caprimulgus europaeus</i>	54	Authors	C <sub>A</sub> , A <sub>A</sub>	Italy	39
European pied flycatcher*	<i>Ficedula hypoleuca</i>	71,75,76	44,49, Authors	C <sub>A</sub> , A <sub>A</sub>	Italy	226
European turtle dove*	<i>Streptopelia turtur</i>	54	Authors	C <sub>A</sub> , A <sub>A</sub>	Italy	29
Garden warbler*	<i>Sylvia borin</i>	54	Authors	C <sub>A</sub> , A <sub>A</sub>	Italy	31
Great reed warbler*	<i>Acrocephalus arundinaceus</i>	54	Authors	C <sub>A</sub> , A <sub>A</sub>	Italy	20
Icterine warbler*	<i>Hippolais icterina</i>	54	Authors	C <sub>A</sub> , A <sub>A</sub>	Italy	29
Northern wheatear*	<i>Oenanthe oenanthe</i>	54	Authors	C <sub>A</sub> , A <sub>A</sub>	Italy	30
Painted bunting*	<i>Passerina ciris</i>	77	77	C <sub>A</sub> , A <sub>A</sub>	USA	60
Sedge warbler*	<i>Acrocephalus schoenobaenus</i>	54	Authors	C <sub>A</sub> , A <sub>A</sub>	Italy	30
Semipalmated plover	<i>Charadrius semipalmatus</i>	69	69	A <sub>A</sub>	Brazil	13
Semipalmated sandpiper	<i>Calidris pusilla</i>	69	69	A <sub>A</sub>	Brazil	14
Spotted flycatcher*	<i>Muscicapa striata</i>	54	Authors	C <sub>A</sub> , A <sub>A</sub>	Italy	29
Spotted sandpiper	<i>Actitis macularius</i>	69	69	A <sub>A</sub>	Brazil	12
Tree pipit*	<i>Anthus trivialis</i>	54,71	44, Authors	C <sub>A</sub> , A <sub>A</sub>	Italy	153
Tree swallow*	<i>Tachycineta bicolor</i>	16,78	46	C <sub>A</sub> , A <sub>A</sub>	Canada	921
Whinchat*	<i>Saxicola rubetra</i>	54,71	44, Authors	C <sub>A</sub> , A <sub>A</sub>	Italy	208
Willow warbler*	<i>Phylloscopus trochilus</i>	54–56	Authors	C <sub>A</sub> , A <sub>A</sub>	Italy	495
Wilson's warbler*	<i>Cardellina pusilla</i>	76	Authors	C <sub>A</sub> , A <sub>A</sub>	USA	102
Wood warbler*	<i>Phylloscopus sibilatrix</i>	54	Authors	C <sub>A</sub> , A <sub>A</sub>	Italy	30
Woodchat shrike*	<i>Lanius senator</i>	54	Authors	C <sub>A</sub> , A <sub>A</sub>	Italy	20
Yellow-legged gull	<i>Larus michahellis</i>	52	52	C <sub>A</sub> , A <sub>A</sub>	Italy	64

**Table 1.** List of species for which published allele data was collected and/or included in the review and data article. Species indicated with an asterisk (\*) were included in the allele dataset for population genetics analyses<sup>34</sup>. The primary study, specific data source, location of the study sites and the sample size (N) is given. C<sub>A</sub>: *Clock* gene alleles, A<sub>A</sub>: *Adcyap1* gene alleles.

of data. The barcode of life data system (BOLD, <https://boldsystems.org/>), which does accept length polymorphism data from microsatellite markers, currently only accepts data for markers used in barcoding or population assignment experiments and does not specifically store data for markers used in behavioural or phenotype associated studies. The European variant archive (EVA, <https://www.ebi.ac.uk/eva/>), which also accepts variant data that includes length polymorphisms, currently only accepts data for species with reference genomes, which is still unavailable for most avian species. To overcome this, we have endeavoured to create a central compilation of the available data in two standard formats which is archived in parallel to this data descriptor; with an additional online version on GitHub<sup>36</sup> (<https://github.com/LSLeClercq/AvianClocksData>) that will be maintained and updated over time as more data is made available. This may greatly facilitate the reuse of the data where it may be applicable to other forms of analyses within migration genetics and beyond.

Common name	Latin binomial	Study	Data	Type	Location	N
African stonechat	<i>Saxicola torquatus</i>	79	79	C <sub>F</sub>	Kenya, Tanzania	172
Asian short-toed lark	<i>Alaudala cheleensis</i>	80	80	C <sub>F</sub>	China	257
Black swan	<i>Cygnus atratus</i>	81	81	Non-CA	Australia	100
Bluethroat	<i>Luscinia svecica</i>	31	31	C <sub>F</sub>	Europe	369
Blue-winged warbler	<i>Vermivora cyanoptera</i>	82	82	Non-CA	USA	24
Canary Island stonechat	<i>Saxicola dacotiae</i>	79	79	C <sub>F</sub>	Canary Islands	61
Chilean swallow	<i>Tachycineta meyeni</i>	16	16	C <sub>F</sub>	Argentina	88
Common buzzard	<i>Buteo buteo</i>	70	48	C <sub>F</sub>	Germany	978
Eurasian blackcap	<i>Sylvia atricapilla</i>	73,74	50,73	C <sub>F</sub>	Europe	936
European roller	<i>Coracias garrulus</i>	14	14	Non-CA	Europe	32
European stonechat	<i>Saxicola rubicola</i>	79	79	C <sub>F</sub>	Europe	382
Golden winged warbler	<i>Vermivora chrysoptera</i>	82	82	Non-CA	USA	42
Great tit	<i>Parus major</i>	83–85	83	C <sub>F</sub>	UK	225
Mangrove swallow	<i>Tachycineta albilinea</i>	16	16	C <sub>F</sub>	Belize	163
Mountain bluebird	<i>Sialia currucoides</i>	18	18	NA	Canada	11
Northern goshawk	<i>Accipiter gentilis</i>	70	48	C <sub>F</sub>	Germany	15
Red kite	<i>Milvus milvus</i>	70	48	C <sub>F</sub> , A <sub>F</sub>	Germany	20
Seychelles warbler	<i>Acrocephalus sechellensis</i>	86	86	Non-CA	Seychelles	57
Siberian stonechat	<i>Saxicola maurus</i>	79	79	C <sub>F</sub>	Kazakhstan, Japan	101
Song sparrow	<i>Melospiza melodia</i>	87	87	Non-CA	Canada	78
Violet-green swallow	<i>Tachycineta thalassina</i>	16	16	C <sub>F</sub>	USA	48
Western bluebird	<i>Sialia mexicana</i>	18	18	NA	Canada	127
White-rumped swallow	<i>Tachycineta leucorrhoa</i>	16	16	C <sub>F</sub>	Argentina	169
Yellow-eyed junco	<i>Junco phaeonotus</i>	72	72	C <sub>F</sub> , A <sub>F</sub>	USA	178

**Table 2.** List of species for which other published data was collected and/or included in the review and data article. The primary study, specific data source, location of the study sites and the sample size (N) is given. C<sub>F</sub>: *Clock* gene frequencies, A<sub>F</sub>: *Adcyap1* gene frequencies, Non-CA: Non clock gene study, NA: Not Available.

## Methods

**Literature search and automated screening.** Literature was searched using systematic review methods, in line with PRISMA Ecology and Evolution guidelines<sup>37</sup>, to identify and synthesize relevant sources. The overall approach is depicted in the PRISMA statement<sup>38</sup> in Fig. 1 that was supplemented with further information on the data retrieval and screening process. Literature was searched between January and September of 2022 on five databases: Scopus (N = 52, [www.scopus.com](http://www.scopus.com)), ScienceDirect (N = 1814, [www.sciencedirect.com](http://www.sciencedirect.com)), Web of Science (N = 140, <https://clarivate.com/>), PubMed (N = 157, <https://pubmed.ncbi.nlm.nih.gov/>), and Dimensions (N = 2746, [www.dimensions.ai](http://www.dimensions.ai)). Databases were searched using an optimized Boolean search string derived from the PICO terms for the aim and objectives of the review. The final search string was as follows: (“Birds” OR “Avian”) AND (“Clock genes” OR “Clock” OR “Adcyap1” OR “Candidate gene”) AND (“Migration” OR “Flying”). As needed, this was complemented by ancillary ‘free term’ searches based on citations in articles or to include other relevant aspects such as “Breeding”, “Moult”, “Genomics”, “Transcriptomics” or “Photoperiod”. For the Scopus and Dimensions database searches, the results were exported in the comma separated value (CSV) format, while the results from the ScienceDirect, Web of Science, and PubMed database search were exported in the research information systems (RIS) format.

Automated screening for inclusion was done through citation network analyses. For the Scopus database, the results were merged and reformatted with the R package ‘Scopus2CitNet 0.1.0.0’ (<https://github.com/MichaelBoireau/Scopus2CitNet>) in RStudio 1.4.1106<sup>39</sup>, running R 4.0.5<sup>40</sup>. The results were subsequently visualized by year in CitNetExplorer 1.0.0., keeping only those papers that overlapped in terms of references cited and the largest connected set (Fig. 2a). The results from the search on the Dimensions and ScienceDirect databases were visualized in VOSviewer 1.6.16<sup>35</sup> by group as well as by year, keeping only those papers that are connected by citations and reference lists (Fig. 2b). The size of bubbles corresponds to citations and the number of cross-links between studies.

**Manual title-abstract screening and full text retrieval.** Sources identified from the citation networks were imported (citation and abstract) into Mendeley citation manager ([www.mendeley.com](http://www.mendeley.com)) for further screening. Several types of studies relating to migration genetics were included in preliminary screening such as candidate gene studies, genomic studies, transcriptomic studies, and epigenetic studies. Studies with a focus on endocrine systems, physiology, or telomeres were excluded. Studies on migration phenology, without an evident genetic link, were also excluded. The inclusion criteria of candidate gene studies were confined to studies that primarily measure *Clock* or *Adcyap1* gene polymorphisms (as well as other candidate genes studied in parallel e.g., *NPAS*, *CREB1*, and *DRD4*: indicated on Fig. 2<sup>34</sup>) within bird populations to compare putative variation to the annual synchronicity in life events and differential migration. These included latitude/longitude/spatial analyses,

Common name	Latin binomial	Study	Data	Type	Location	N
American redstart*	<i>Setophaga ruticilla</i>	34	Authors	C <sub>A</sub> , A <sub>A</sub>	Canada	26
Barolo shearwater	<i>Puffinus baroli</i>	88	Authors	C <sub>A</sub>	Portugal	15
Boyd's shearwater	<i>Puffinus boydi</i>	88	Authors	C <sub>A</sub>	Cape Verde	25
Common chiffchaff*	<i>Phylloscopus collybita</i>	34	Authors	C <sub>A</sub>	Sweden, Kazakhstan	55
Common yellowthroat*	<i>Geothlypis trichas</i>	34	Authors	C <sub>A</sub> , A <sub>A</sub>	Canada	31
Great shearwater	<i>Ardenna gravis</i>	88	Authors	C <sub>A</sub>	UK	25
Hermit thrush*	<i>Catharus guttatus</i>	34	Authors	C <sub>A</sub> , A <sub>A</sub>	Canada	30
Magnolia warbler*	<i>Setophaga magnolia</i>	34	Authors	C <sub>A</sub> , A <sub>A</sub>	Canada	33
Manx shearwater	<i>Puffinus puffinus</i>	88	Authors	C <sub>A</sub>	Iceland	23
Swainson's thrush*	<i>Catharus ustulatus</i>	34	Authors	C <sub>A</sub> , A <sub>A</sub>	Canada	29
White-throated sparrow*	<i>Zonotrichia albicollis</i>	34	Authors	C <sub>A</sub> , A <sub>A</sub>	Canada	32
Yelkouan shearwater	<i>Puffinus yelkouan</i>	88	Authors	C <sub>A</sub>	France	15

**Table 3.** List of species for which unpublished data was collected and/or included in the review and data article. Species indicated with an asterisk (\*) were included in the allele dataset for population genetics analyses<sup>34</sup>. The primary study, specific data source, location of the study sites and the sample size (N) is given. C<sub>A</sub>: *Clock* gene alleles, A<sub>A</sub>: *Adcyap1* gene alleles.

timing of migration, migratory restlessness, timing of egg laying/breeding, clutch size, moult, urbanisation, and exploratory behaviour. The final set of studies that passed preliminary screening were sought during full text retrieval and added to the imported reference if it wasn't already included. A total of 66 studies were included in the final review of which 34 were candidate gene studies and 32 were other, migration related, studies using genetic methods. Some basic scientometric assessments of the final set of studies, including the plotting of publications per year (Fig. 3) as well as the geographic distribution (Fig. 4) of studies, was conducted using ABCal version 1.0.2<sup>41</sup> (<https://github.com/LSLeClercq/ABCal>).

**Published datasets.** A total of 34 studies were identified that used a candidate gene approach for which data retrieval was done. Data was retrieved from either the main text, supplementary material of the article, online data repositories such as Dryad<sup>42–49</sup> and Figshare<sup>50</sup>, or additional data received directly from authors. Data types varied from allele frequencies to individual level diploid allele data. Allele data for the Barn swallow<sup>51</sup> was retrieved from the text while data for the Yellow-legged gull<sup>52</sup> was extracted from images using WebPlotDigitizer version 4.6<sup>53</sup>. Allele data was generally derived from a single source with the exception of the European pied flycatcher<sup>44,49</sup> and Willow warbler<sup>54–56</sup>. The species, data sources, and data types are summarized in Table 1 along with the sampling location and sample sizes. Frequency data was available for most published studies, with the exception of the bluebird species<sup>18</sup>, and those species for which allele data was unavailable are summarised in Table 2. This includes species for which only frequency data was reported, species for which a non-clock gene approach was used, and studies for which only data summaries without frequencies were reported.

**Unpublished datasets.** This study included unpublished data for twelve species in total, summarised in Table 3. The six North American species were sampled at Long Point Old Cut, Ontario, Canada, and included the American redstart (N = 26), Common yellowthroat (N = 31), Hermit thrush (N = 30), Magnolia warbler (N = 33), Swainson's thrush (N = 29), and White-throated sparrow (N = 32). The six European species included the Common chiffchaff (N = 55) and five species of shearwaters: Barolo shearwater (N = 15), Boyd's shearwater (N = 25), Great shearwater (N = 25), Manx shearwater (N = 23), and Yelkouan shearwater (N = 15). The Common chiffchaff was sampled from several locations in Sweden (N = 30, subspecies *abietinus*) and Kazakhstan (N = 25, spp. *tristis*). Blood samples were taken from the brachial vein and stored in SET buffer at –80 °C. Shearwaters were sampled from several locations in Europe including France and Portugal while several species were sampled from islands such as Iceland, Cape Verde, and territories of the United Kingdom such as Gough Island. A 1 ml blood sample was taken from the tarsal or the brachial vein during geolocator retrieval. Samples were collected in 1.5 ml plastic tubes containing 70% ethanol and stored at –20 °C until further analysis.

Samples were genotyped using established methods<sup>54</sup>. Briefly, samples of North American species were preserved in a buffer at room temperature until extraction with the ArchivePure DNA purification kit (5 PRIME, Hilden, Germany). Then, polymorphism at *Clock* and *Adcyap1* 3'-UTR was determined as before<sup>54</sup>, with PCR products labelled with HEX (*Clock*), 6-FAM (*Clock* and *Adcyap1*) or TAMRA (*Adcyap1*) dyes. For the Common chiffchaff, genomic DNA was extracted using a standard ammonium acetate protocol. All 55 samples were successfully genotyped and analysed for length polymorphism in the poly-Q repeat of the *Clock* gene following previously published protocols<sup>31</sup>. For Shearwater samples, total genomic DNA was extracted from blood samples using the Speedtools® Tissue DNA Extraction kit (Biotools, Madrid, Spain) following the manufacturer's instructions. Genotyping was subsequently performed with methods adapted from the literature<sup>31</sup>. Briefly, PCR products were generated with shearwater specific primers for the *Clock* gene labelled with 6-FAM or HEX, followed by fragment analysis as in<sup>54</sup> to determine the size of the poly-Q repeat.



Field name	Data
<i>General (Index):</i>	
Species	Common name in English for species
Clock	Logical binary for data availability of <i>Clock</i> gene e.g., “Yes” or “No”
<i>Adcyap1</i>	Logical binary for data availability of <i>Adcyap1</i> gene e.g., “Yes” or “No”
Code	Abbreviation used for species tabs
Sample (N)	Size (N) of the total individuals for which data are available
<i>Taxonomy (Index):</i>	
Genus	Latin name for genus e.g., “Hirundo”
Species	Latin name for species e.g., “rustica”
Family	Latin name for family e.g., “Hirundinidae”
Superfamily	Latin name for superfamily e.g., “Locustelloidea”
Parvorder	Latin name for parvorder e.g., “Sylviida”
Order	Latin name for order e.g., “Passeriformes”
<i>Species sheet:</i>	
Species	Common name in English for species
Sample ID	Sample ID used in raw data for individuals
Clock 1	1 <sup>st</sup> diploid allele for <i>Clock</i> gene (individual) as $Q_N$
Clock 2	2 <sup>nd</sup> diploid allele for <i>Clock</i> gene (individual) as $Q_N$
Sum	Sum of two alleles for <i>Clock</i> gene as $Q_N$
Mean	Mean value of diploid alleles for <i>Clock</i> gene as $Q_N$
<i>Adcyap 1</i>	1 <sup>st</sup> diploid allele for <i>Adcyap1</i> gene (individual) in base pairs (bp)
<i>Adcyap 2</i>	2 <sup>nd</sup> diploid allele for <i>Adcyap1</i> gene (individual) in base pairs (bp)
Sum	Sum of two alleles for <i>Adcyap1</i> gene in base pairs (bp)
Mean	Mean value of diploid alleles for <i>Adcyap1</i> gene in base pairs (bp)

**Table 4.** Description of field names and data for workbook and CSV files.

## Data Records

The data collated during the systematic review and meta-analysis were made available to via the Zenodo repository at the time of publication. Additional inclusion and exclusion criteria were applied and a final set of 40 species (indicated by asterisk in Tables 1, 3) were included in the comparative analyses using mantel and phylogenetic generalised least squares methods to test for an association between migratory phenotypes and candidate gene genotypes<sup>34,57</sup>. This data are available on Zenodo<sup>57</sup>, and includes a workbook with the allele data as well as a results workbook with various population genetics measures including allele frequencies, Homozygosity ( $H_o$ ), Heterozygosity ( $H_e$ ), Hardy-Weinberg equilibrium<sup>58,59</sup>, and Ewens-Watterson<sup>60</sup> results. The complete dataset was reformatted for distribution with this data descriptor and is available from two sources, from the Figshare<sup>61</sup> depository, as submitted with this article, and from a maintained repository with version histories on GitHub<sup>36</sup>.

Data (version 1.0.2) are available as a spreadsheet workbook, labelled “Avian Clock Gene Dataset” with multiple sheets. The first sheet of the workbook, labelled “Index”, contains the table of contents which has several columns (Table 4) that list species by common names, indicates data availability for *Clock* and *Adcyap1*, and total sample size (N). Furthermore, the taxonomic classifications including genus, species, family, superfamily, parvorder, and order are also given. The species codes are hyperlinked to the allele data for individual species, contained in separate sheets within the same workbook. Individual sheets for species contain several columns including the species name, sample ID, and diploid alleles for *Clock* and/or *Adcyap1* genes. Alleles are expressed as the number of polyglutamine repeats ( $Q_N$ ) for *Clock* while the *Adcyap1* alleles represent the amplified fragment length in base pairs (bp). The sum and average of alleles is also provided, and missing data is labelled as NA. For the purpose of individual species analyses, the species sheets from the workbook are also provided as individual comma separated value (CSV) files. The same data is also available on GitHub with the workbooks available in the root directory while the individual CSV files are available in a subfolder with the title “CSV”. The repository also contains a “README” file which provides some basic background and details on the data. Both the workbook as well as CSV files can be read by Microsoft® Office (<https://www.office.com/>) as well as StarOffice™ (<https://www.staroffice.com/>), OpenOffice™ (<https://www.openoffice.org/>), and LibreOffice™ (<https://www.libreoffice.org/>).

## Technical Validation

Allele data comprises the heterozygous or homozygous diploid allele for one or both studied clock genes as well as the sum and average of allele sizes. The data for *Clock* was normalized according to the poly-glutamine repeat size ( $Q_N$ ) by subtracting the conserved non-repeat size ( $L_C$ ) in base pairs from the total fragment size ( $L_T$ ) and dividing by codon size, following Eq. 1.

$$Q_N = (L_T - L_C)/3 \quad (1)$$

Data for *Adcyap1* was generated using the same published primers and was kept as the total fragment size.

## Code availability

The custom R code used to convert data retrieved from Scopus to the appropriate format for visualisation in CitNetExplorer is available from GitHub (<https://github.com/MichaelBoireau/Scopus2CitNet>). The custom PYTHON script used for plotting the scientometric aspects of the included literature is also available from GitHub (<https://github.com/LSLeClercq/ABCAL>).

Received: 7 July 2023; Accepted: 1 November 2023;

Published: 9 November 2023

## References

1. Leclerc, B. *et al.* Photoperiodic modulation of clock gene expression in the avian pre-mammillary nucleus. *J. Neuroendocrinol.* **22**, 119–128 (2010).
2. Pancerasa, M., Ambrosini, R., Saino, N. & Casagrandi, R. Barn swallows long-distance migration occurs between significantly temperature-correlated areas. *Sci. Rep.* **8**, 12359 (2018).
3. Saino, N. *et al.* Temperature and rainfall anomalies in Africa predict timing of spring migration in trans-Saharan migratory birds. *Clim. Res.* **35**, 123–134 (2007).
4. Studds, C. E. & Marra, P. P. Linking fluctuations in rainfall to nonbreeding season performance in a long-distance migratory bird, *Setophaga ruticilla*. *Clim. Res.* **35**, 115–122 (2007).
5. Hau, M. & Gwinner, E. Food as a circadian zeitgeber for house sparrows: The effect of different food access durations. *J. Biol. Rhythms* **11**, 196–207 (1996).
6. Sockman, K. W. & Hurlbert, A. H. How the effects of latitude on daylight availability may have influenced the evolution of migration and photoperiodism. *Funct. Ecol.* **34**, 1752–1766 (2020).
7. Norevik, G., Åkesson, S., Andersson, A., Bäckman, J. & Hedenström, A. The lunar cycle drives migration of a nocturnal bird. *PLoS Biol.* **17**, e3000456 (2019).
8. Pyle, P., Nur, N., Henderson, R. P. & DeSante, D. F. The Effects of Weather and Lunar Cycle on Nocturnal Migration of Landbirds at Southeast Farallon Island, California. *Condor* **95**, 343 (1993).
9. Roques, S. *et al.* When to depart from a stopover site? Time since arrival matters more than current weather conditions. *Ornithology* **139**, 1–13 (2022).
10. Aguilar-Roblero, R., Díaz-Muñoz, M. & Fanjul-Moles, M. L. *Mechanisms of circadian systems in animals and their clinical relevance. Mechanisms of Circadian Systems in Animals and Their Clinical Relevance.* <https://doi.org/10.1007/978-3-319-08945-4> (Springer International Publishing, 2015).
11. Fry, H., Kirwan, G. M. & Boesman, P. F. D. Lilac-breasted Roller (*Coracias caudatus*). in *Birds of the World* (eds. del Hoyo, J., Elliott, A., Sargatal, J., Christie, D. & de Juana, E.) <https://doi.org/10.2173/bow.librol2.01.1> (Cornell Lab of Ornithology, 2021).
12. Woodall, P. F. Woodland Kingfisher (*Halcyon senegalensis*), version 1.0. in *Birds of the World* (eds. *et al.*) <https://doi.org/10.2173/BOW.WOOKIN1.01> (Cornell Lab of Ornithology, 2020).
13. Fry, H. & Boesman, P. F. D. European Bee-eater (*Merops apiaster*), version 1.0. in *Birds of the World* (eds. del Hoyo, J., Elliott, A., Sargatal, J., Christie, D. A. & de Juana, E.) <https://doi.org/10.2173/BOWEUBAT1.01> (Cornell Lab of Ornithology, 2020).
14. Finch, T. *et al.* A pan-European, multipopulation assessment of migratory connectivity in a near-threatened migrant bird. *Divers. Distrib.* **21**, 1051–1062 (2015).
15. Mwale, M. *et al.* Genetic and morphological variation of Woodland Kingfisher *Halcyon senegalensis* reveals cryptic mitochondrial lineages and patterns of mitochondrial–nuclear discordance. *Ostrich* **93**, 192–207 (2022).
16. Dor, R. *et al.* Clock gene variation in tachycineta swallows. *Ecol. Evol.* **2**, 95–105 (2012).
17. Hegemann, A., Fudickar, A. M. & Nilsson, J. Å. A physiological perspective on the ecology and evolution of partial migration. *J. Ornithol.* **2019** **160**, 893–905 (2019).
18. Sauve, D., Dale, C. A., Tigano, A., Ratcliffe, L. M. & Friesen, V. L. Do candidate genes for migration and behavior explain migratory variation in bluebirds (*Sialia* spp.)? *Wilson J. Ornithol.* **132**, 820–829 (2021).
19. Carey, C. The impacts of climate change on the annual cycles of birds. *Philos. Trans. R. Soc. B Biol. Sci.* **364**, 3321–3330 (2009).
20. Saino, N. *et al.* Climate change effects on migration phenology may mismatch brood parasitic cuckoos and their hosts. *Biol. Lett.* **5**, 539–541 (2009).
21. Beresford, A. E. *et al.* Phenology and climate change in Africa and the decline of Afro-Palaearctic migratory bird populations. *Remote Sens. Ecol. Conserv.* **5**(1), 55–69 (2019).
22. Mueller, J. C., Partecke, J., Hatchwell, B. J., Gaston, K. J. & Evans, K. L. Candidate gene polymorphisms for behavioural adaptations during urbanization in blackbirds. *Mol. Ecol.* **22**, 3629–3637 (2013).
23. Güneralp, B., Lwasa, S., Masundire, H., Parnell, S. & Seto, K. C. Urbanization in Africa: Challenges and opportunities for conservation. *Environ. Res. Lett.* **13**, 015002 (2018).
24. Navarro, L. M. & Pereira, H. M. Rewilding abandoned landscapes in Europe. in *Rewilding European Landscapes* 3–23. [https://doi.org/10.1007/978-3-319-12039-3\\_1](https://doi.org/10.1007/978-3-319-12039-3_1) (2015).
25. Hoogendoorn, G., Meintjes, D., Kelso, C. & Fitchett, J. Tourism as an incentive for rewilding: the conversion from cattle to game farms in Limpopo province, South Africa. *J. Ecotourism* **18**, 309–315 (2019).
26. Delmore, K. E. *et al.* Genomic analysis of a migratory divide reveals candidate genes for migration and implicates selective sweeps in generating islands of differentiation. *Mol. Ecol.* **24**, 1873–1888 (2015).
27. Saino, N. *et al.* Inter-generational resemblance of methylation levels at circadian genes and associations with phenology in the barn swallow. *Sci. Rep.* **9**, 1–16 (2019).
28. Frias-Soler, R. C., Pildain, L. V., Párra, L. G., Wink, M. & Bairlein, F. Transcriptome signatures in the brain of a migratory songbird. *Comp. Biochem. Physiol. - Part D Genomics Proteomics* **34**, 100681 (2020).
29. Sharma, A., Tripathi, V. & Kumar, V. Control and adaptability of seasonal changes in behavior and physiology of latitudinal avian migrants: Insights from laboratory studies in Palaearctic-Indian migratory buntings. *J. Exp. Zool. Part A Ecol. Integr. Physiol.* **337**, 902–918 (2022).
30. Åkesson, S. *et al.* Timing avian long-distance migration: from internal clock mechanisms to global flights. *Philos. Trans. R. Soc. B Biol. Sci.* **372**, 20160252 (2017).
31. Johnsen, A. *et al.* Avian Clock gene polymorphism: Evidence for a latitudinal cline in allele frequencies. *Mol. Ecol.* **16**, 4867–4880 (2007).
32. Steinmeyer, C., Mueller, J. C. & Kempnaers, B. Search for informative polymorphisms in candidate genes: Clock genes and circadian behaviour in blue tits. *Genetica* **136**, 109–117 (2009).
33. Steinmeyer, C., Kempnaers, B. & Mueller, J. C. Testing for associations between candidate genes for circadian rhythms and individual variation in sleep behaviour in blue tits. *Genetica* **140**, 219–228 (2012).
34. Le Clercq, L. *et al.* Time trees and clock genes: a systematic review and comparative analysis of contemporary avian migration genetics. *Biol. Rev.* **98**, 1051–1080 (2023).

35. van Eck, N. J. & Waltman, L. Citation-based clustering of publications using CitNetExplorer and VOSviewer. *Scientometrics* **111**, 1053–1070 (2017).
36. Le Clercq, L. S. AvianClocksData: Dataset of Clock and Adcyap1 alleles for Birds. Version 1.0.2. <https://github.com/LSLeClercq/AvianClocksData> (2023).
37. O’Dea, R. E. *et al.* Preferred reporting items for systematic reviews and meta-analyses in ecology and evolutionary biology: a PRISMA extension. *Biol. Rev.* **96**, 1695–1722 (2021).
38. Haddaway, N. R., Page, M. J., Pritchard, C. C. & McGuinness, L. A. PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. *Campbell Syst. Rev.* **18**, e1230 (2022).
39. RStudio Team. RStudio: Integrated Development Environment for R. Version 4.0.5. <http://www.rstudio.com/> (2021).
40. R Core Team. R: A Language and Environment for Statistical Computing. Version 1.4.1106. <https://www.r-project.org/> (2020).
41. Le Clercq, L. S. ABCal: Author Bias Computation and Scientometric Plotting. Version 1.0.2. <https://github.com/LSLeClercq/ABCal> (2023).
42. Liedvogel, M., Cornwallis, C. K. & Sheldon, B. C. Data from: Integrating candidate gene and quantitative genetic approaches to understand variation in timing of breeding in wild tit populations. *Dryad Dataset*. <https://doi.org/10.5061/dryad.6n3d6109> (2012).
43. Ralston, J. *et al.* Data from: Length polymorphisms at two candidate genes explain variation of migratory behaviors in blackpoll warblers (*Setophaga striata*). *Dryad Dataset*. <https://doi.org/10.5061/dryad.d10qb58> (2019).
44. Saino, N. *et al.* Data from: Polymorphism at the Clock gene predicts phenology of long-distance migration in birds. *Dryad Dataset* <https://doi.org/10.5061/dryad.j22j1> (2015).
45. Liedvogel, M., Szulkin, M., Knowles, S., Wood, M. & Sheldon, B. Data from: Phenotypic correlates of Clock gene variation in a wild blue tit population: evidence for a role in seasonal timing of reproduction. *Dryad Dataset*. <https://doi.org/10.5061/dryad.1303> (2010).
46. Bourret, A. & Garant, D. Data from: Candidate gene-environment interactions and their relationships with timing of breeding in a wild bird population. *Dryad Dataset* <https://doi.org/10.5061/dryad.f7t25> (2015).
47. Krist, M., Munclinger, P., Briedis, M. & Adamík, P. Data from: The genetic regulation of avian migration timing: combining candidate genes and quantitative genetic approaches in a long-distance migrant. *Dryad Dataset*. <https://doi.org/10.5061/dryad.gb5mkkwpr> (2021).
48. Chakarov, N., Jonker, R. M., Boerner, M., Hoffman, J. I. & Krüger, O. Data from: Variation at phenological candidate genes correlates with timing of dispersal and plumage morph in a sedentary bird of prey. *Dryad Dataset* <https://doi.org/10.5061/dryad.r058p> (2013).
49. Kuhn, K. *et al.* Data from: Differentiation in neutral genes and a candidate gene in the pied flycatcher: using biological archives to track global climate change. *Dryad Dataset* <https://doi.org/10.5061/dryad.6dc6k> (2014).
50. Mettler, R., Segelbacher, G. & Martin Schaefer, H. Fileset: PONE-D-15-12289. *Figshare Dataset*. <https://doi.org/10.6084/m9.figshare.1492930.v1> (2016).
51. Bazzi, G. *et al.* Clock gene polymorphism and scheduling of migration: A geolocator study of the barn swallow *Hirundo rustica*. *Sci. Rep.* **5**, 12443 (2015).
52. Romano, A. *et al.* Circadian genes polymorphism and breeding phenology in a resident bird, the yellow-legged gull. *J. Zool.* **304**, 117–123 (2018).
53. Rohatgi, A. WebPlotDigitizer: Version 4.6. <https://automeris.io/WebPlotDigitizer> (2022).
54. Bazzi, G. *et al.* Clock gene polymorphism, migratory behaviour and geographic distribution: a comparative study of trans-Saharan migratory birds. *Mol. Ecol.* **25**, 6077–6091 (2016).
55. Bazzi, G. *et al.* Candidate genes have sex-specific effects on timing of spring migration and moult speed in a long-distance migratory bird. *Curr. Zool.* **63**, 479–486 (2017).
56. Sokolovskis, K. *et al.* Phenotypic and genetic characterization of the East Siberian Willow Warbler (*Phylloscopus trochilus yakutensis* Ticehurst, 1935) in relation to the European subspecies. *J. Ornithol.* **160**, 721–731 (2019).
57. Le Clercq, L. S. *et al.* Time trees and Clock genes: a Systematic Review and Comparative Analysis of Contemporary Avian Migration Genetics (Dataset). *Zenodo*. <https://doi.org/10.5281/ZENODO.6637839> (2022).
58. Hardy, G. H. Mendelian proportions in a mixed population. *Science*. **28**, 49–50 (1908).
59. Weinberg, W. On the demonstration of heredity in man. *Pap. Hum. Genet.* (1963).
60. Slatkin, M. An exact test for neutrality based on the Ewens sampling distribution. *Genet. Res.* **64**, 71–74 (1994).
61. Le Clercq, L. S. *et al.* Birds of a feather flock together—A compilation of data for two candidate genes for migration genetics. *Figshare Dataset*. <https://doi.org/10.6084/m9.figshare.23623935.v1> (2023).
62. Dor, R. *et al.* Low variation in the polymorphic Clock gene poly-Q region despite population genetic structure across Barn Swallow (*Hirundo rustica*) populations. *PLoS One* **6**, e0028843 (2011).
63. Caprioli, M. *et al.* Clock gene variation is associated with breeding phenology and maybe under directional selection in the migratory barn swallow. *PLoS One* **7**, e0035140 (2012).
64. Parody-Merino, Á. M., Battley, P. F., Conklin, J. R. & Fidler, A. E. No evidence for an association between Clock gene allelic variation and migration timing in a long-distance migratory shorebird (*Limosa lapponica baueri*). *Oecologia* **191**, 843–859 (2019).
65. Ralston, J. *et al.* Length polymorphisms at two candidate genes explain variation of migratory behaviors in blackpoll warblers (*Setophaga striata*). *Ecol. Evol.* **9**, 8840–8855 (2019).
66. Liedvogel, M., Szulkin, M., Knowles, S. C. L., Wood, M. J. & Sheldon, B. C. Phenotypic correlates of clock gene variation in a wild blue tit population: Evidence for a role in seasonal timing of reproduction. *Mol. Ecol.* **18**, 2444–2456 (2009).
67. Liedvogel, M., Cornwallis, C. K. & Sheldon, B. C. Integrating candidate gene and quantitative genetic approaches to understand variation in timing of breeding in wild tit populations. *J. Evol. Biol.* **25**, 813–823 (2012).
68. Krist, M., Munclinger, P., Briedis, M. & Adamík, P. The genetic regulation of avian migration timing: combining candidate genes and quantitative genetic approaches in a long-distance migrant. *Oecologia* **196**, 373–387 (2021).
69. de Almeida Miranda, D. *et al.* Shorebirds’ Longer Migratory Distances Are Associated With Larger ADCYAP1 Microsatellites and Greater Morphological Complexity of Hippocampal Astrocytes. *Front. Psychol.* **12**, 6679 (2022).
70. Chakarov, N., Jonker, R. M., Boerner, M., Hoffman, J. I. & Krüger, O. Variation at phenological candidate genes correlates with timing of dispersal and plumage morph in a sedentary bird of prey. *Mol. Ecol.* **22**, 5430–5440 (2013).
71. Saino, N. *et al.* Polymorphism at the Clock gene predicts phenology of long-distance migration in birds. *Mol. Ecol.* **24**, 1758–1773 (2015).
72. Peterson, M. P. *et al.* Variation in candidate genes CLOCK and ADCYAP1 does not consistently predict differences in migratory behavior in the songbird genus Junco. *F1000Research* **2**, 115 (2013).
73. Mueller, J. C., Pulido, F. & Kempenaers, B. Identification of a gene associated with avian migratory behaviour. *Proc. R. Soc. B Biol. Sci.* **278**, 2848–2856 (2011).
74. Mettler, R., Segelbacher, G. & Schaefer, H. M. Interactions between a Candidate Gene for Migration (ADCYAP1), Morphology and Sex Predict Spring Arrival in Blackcap Populations. *PLoS One* **10**, e0144587 (2015).
75. Kuhn, K. *et al.* Differentiation in neutral genes and a candidate gene in the pied flycatcher: Using biological archives to track global climate change. *Ecol. Evol.* **3**, 4799–4814 (2013).
76. Bazzi, G. *et al.* Adcyap1 polymorphism covaries with breeding latitude in a Nearctic migratory songbird, the Wilson’s warbler (*Cardellina pusilla*). *Ecol. Evol.* **6**, 3226–3239 (2016).

77. Contina, A., Bridge, E. S., Ross, J. D., Shipley, J. R. & Kelly, J. F. Examination of *Clock* and *Adcyap1* gene variation in a neotropical migratory passerine. *PLoS One* **13**, e0190859 (2018).
78. Bourret, A. & Garant, D. Candidate gene-environment interactions and their relationships with timing of breeding in a wild bird population. *Ecol. Evol.* **5**, 3628–3641 (2015).
79. Justen, H. *et al.* Population-specific association of *Clock* gene polymorphism with annual cycle timing in stonechats. *Sci. Rep.* **12**, 1–13 (2022).
80. Zhang, S., Xu, X., Wang, W., Yang, W. & Liang, W. *Clock* gene is associated with individual variation in the activation of reproductive endocrine and behavior of Asian short toed lark. *Sci. Rep.* **7**, 1–8 (2017).
81. Van Dongen, W. F. D., Robinson, R. W., Weston, M. A., Mulder, R. A. & Guay, P. J. Variation at the *DRD4* locus is associated with wariness and local site selection in urban black swans. *BMC Evol. Biol.* **15**, 253 (2015).
82. Toews, D. P. L., Taylor, S. A., Streby, H. M., Kramer, G. R. & Lovette, I. J. Selection on *VPS13A* linked to migration in a songbird. *Proc. Natl. Acad. Sci. USA* **116**, 18272–18274 (2019).
83. Liedvogel, M. & Sheldon, B. C. Low variability and absence of phenotypic correlates of *Clock* gene variation in a great tit *Parus major* population. *J. Avian Biol.* **41**, 543–550 (2010).
84. Korsten, P. *et al.* Association between *DRD4* gene polymorphism and personality variation in great tits: A test across four wild populations. *Mol. Ecol.* **19**, 832–843 (2010).
85. Mueller, J. C. *et al.* Haplotype structure, adaptive history and associations with exploratory behaviour of the *DRD4* gene region in four great tit (*Parus major*) populations. *Mol. Ecol.* **22**, 2797–2809 (2013).
86. Edwards, H. A., Hajduk, G. K., Durieux, G., Burke, T. & Dugdale, H. L. No association between personality and candidate gene polymorphisms in a wild bird population. *PLoS One* **10**, e0138439 (2015).
87. Posliff, C. D. Correlations in movement behaviour over large and small geographic scales in song sparrows (*Melospiza melodia*). (Western University of Ontario, 2020).
88. Ferrer Obiol, J. New insights into the genetic base of bird migration: a population genetics study based on candidate genes. (University of Barcelona, 2015).

## Acknowledgements

Open access provided by the library of University of the Free State (Grant number: 13560591). The authors would like to thank Chris Guglielmo, Yolanda Morbey, and Ivan Maggini for support with field data collection at Long Point Bird Observatory (Canada), Staffan Bensch for field data collection in both Sweden and Kazakhstan, and Andrea Galimberti, Cristina Possenti, Fernando Spina and the late Nicola Saino for support with genetic analyses and field data collection in Ventotene Island (Italy). Images were created/edited in BioRender.com. This work is based on the research supported wholly/in part by the National Research Foundation of South Africa (Grant Number: 112062). The study received partial financial support from the PRIN 2017 funding scheme (grant number 20178T2PSW to Diego Rubolini), from the ICREA Academia award (to Jacob González-Solís), and from the Ministerio de Ciencia e Innovación, Agencia Estatal de Investigación, España (PID2021-125792NB-I00 to Marta Riutort).

## Author contributions

Le Clercq, L.S.: Conceptualisation, Methodology, Data Curation, Formal Analysis, Writing – Original Draft, Writing – Review & Editing, Funding acquisition. Bazzi, G.: Resources, Data Curation. Ferrer Obiol, J.: Resources, Data Curation, Writing – Review & Editing. Cecere, J.G.: Resources, Data Curation, Funding acquisition. Gianfranceschi, L.: Resources, Data Curation. Grobler, J.P.: Writing – Review & Editing, Resources, Supervision. Kotzé, A.: Writing – Review & Editing, Resources, Supervision. Riutort León, M.: Resources, Data Curation. González-Solís, J.: Resources, Data Curation. Rubolini, D.: Resources, Data Curation, Writing – Review & Editing, Funding acquisition. Liedvogel, M.: Data Curation, Writing – Review & Editing. Dalton, D.L.: Conceptualisation, Writing – Review & Editing, Resources, Supervision.

## Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Additional information

**Correspondence** and requests for materials should be addressed to L.-S.L.C.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023, corrected publication 2024