



OPEN

DATA DESCRIPTOR

# A dataset of diversity and distribution of rodents and shrews in China

Jin-Jin Chen<sup>1</sup>, Qiang Xu<sup>1</sup>, Tao Wang<sup>1</sup>, Fan-Fei Meng<sup>1</sup>, Zhi-Wei Li<sup>1,2</sup>, Li-Qun Fang<sup>1</sup> & Wei Liu<sup>1</sup>

The rodents and shrews are important reservoirs for a large number of zoonotic pathogens. Here by performing a literature review, we determined the occurrence and distribution of rodents and shrews in China at three scales including province, city, and county levels. The peer-reviewed papers published in English and Chinese were collected, standard procedures were applied in reference books, field surveys and websites to remove duplicates, and information on recorded locations of rodents and shrews was extracted. The dataset contains 13,911 records of geo-referenced occurrences for 364 rodents and shrews distributed over 1,663 locations distinguished. As pathogens continue to emerge from rodents and shrews, this dataset could assist efforts to put preliminary bounds around a variety of spatial analyses of rodents and shrews, facilitate a better understanding of the transmission risk of the pathogens they carry, and be helpful for assessing the risk of future emergence of rodent-borne zoonoses.

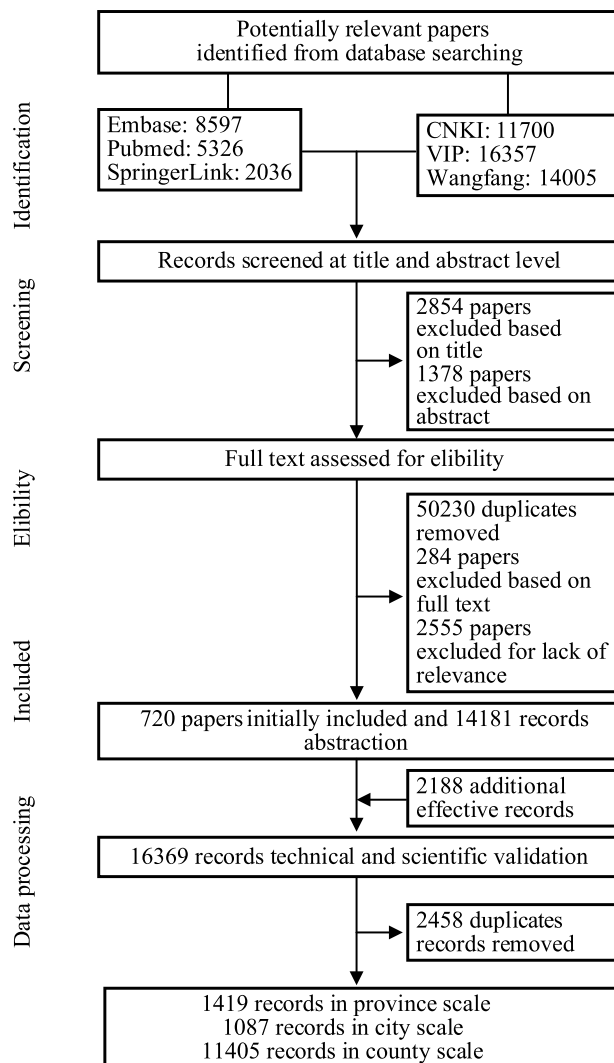
## Background & Summary

In the last decades, the incidence of human diseases caused by zoonotic viruses, bacteria, and parasites that are associated with small mammal reservoirs appears to have increased<sup>1</sup>. Rodents are the most abundant and diversified order of mammals, comprising approximately 42% of global mammalian biodiversity<sup>2</sup>, with an extensive range of species distributed in all continents except Antarctica<sup>3</sup>. Rodents and shrews include some small mammals, such as EULIPOTYPHILA, LAGOMORPHA, RODENTIA and SCANDENTIA<sup>4-6</sup>. They live in close contact with human beings, acting as a nexus among humans, domestic animals, pets, arthropod vectors (ticks, mites, fleas), and other wildlife, thus creating opportunity for rodents to transmit pathogens to humans, with hantaviruses, lassa viruses and *Yersinia pestis* as the prominent examples. Moreover, rodents can help to maintain pathogen transmission cycles in a number of different environments, varying from densely populated urban areas to rural areas and in the wilderness<sup>7</sup>. With an increased incidence and lack of effective prophylaxis or vaccines for most rodent-borne illnesses, a substantial burden remained on human health and agricultural economics in various countries<sup>8</sup>.

China is a country that hosts diverse climatic and ecosystems suitable for the propagation of a high diversity of rodents and shrews<sup>9</sup>. During the last decades we have seen a rise in human diseases that are associated with small-mammal reservoirs. There was relevant literature on rodent and rodent-borne pathogens in China<sup>10</sup>. However, the lately implemented policies that aim to protect, restore and conserve biodiversity in China, had resulted in ecologic changes promoting rapid increases in populations of rodents and shrews<sup>11</sup>. Other human activities and behaviors, including changing human settlement patterns (especially rapid urbanization in developed regions), the widespread migration of the human population, intensive trade activities, and explosive travel, are also enhancing the expanded transmission to human in new regions<sup>12-14</sup>. There is a great need to understand current knowledge on the diversity and distribution of rodents and shrews, so to attain a better understanding of their contribution to disease and a quantification of the risks per disease. This information is important for clinicians, researchers and professionals in the field.

The goal of this systematic review of published literature was to assembly a comprehensive database on the reports of contemporary occurrence and distribution of rodents and shrews in the mainland of China, and aggregate those indicators into a single open source dataset. We created a public database for rodents and shrews

<sup>1</sup>State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing, P. R. China. <sup>2</sup>Department of Epidemiology and Health Statistics, School of Public Health, Capital Medical University, Beijing, P. R. China. ✉e-mail: [fang\\_lq@163.com](mailto:fang_lq@163.com); [liuwei@bmi.ac.cn](mailto:liuwei@bmi.ac.cn)



**Fig. 1** Flow chart of retrieval, screening, and inclusion of systematic database.

abundance and prevalence in China, reported from 1950 to 2021, encompassing 364 species of rodents and shrews in 1,660 counties and districts with 13,911 records. As rodent and shrew species-related records were better documented in Chinese, a study breaking language barriers for a more complete picture of spatial diversity in China and against a true baseline estimate of the geographic distribution of rodents and shrews would be useful for the new disease threats associated with rodent-borne zoonoses.

## Methods

**Data collection.** The protocols stated here have been adapted from previously published literature extraction efforts. A guide to our extraction has been included in Fig. 1, and shows the overarching process we followed to produce this dataset. We searched PubMed and ISI Web of Science for literature published in English, and China National Knowledge Infrastructure (CNKI), VIP database and Wanfang database as the core sources published in Chinese through the duration from 1950 to 2021. We used the following search terms: “Rodent”, “Mouse”, “Rat”, “Shrew”, and “China”, in combination with each of the 18 families of rodent and shrew species (either in English or in Chinese). A secondary manual search of the references cited in these articles was also performed to find relevant articles. For supplementation, the historical collection of rodent-related records that were documented in Chinese publications, with the multiple books (Fauna sinica, mammal, vol.6(2), cricetidae<sup>15</sup>, A guide to the mammals of China<sup>16</sup>, Distribution of mammal species in China<sup>17</sup>, Colored atlas of Chinese mammalians<sup>18</sup>, Mammalians of Tibet<sup>19</sup>, China’s red list of biodiversity: vertebrates, volume I, mammal<sup>20</sup>, Atlas of epidemiology of natural focus diseases in China<sup>21</sup>, Fauna of Beijing<sup>22</sup>, Fauna of Guizhou<sup>23</sup>, etc.), were also reviewed. Moreover, we standardize the species lists of the mammals of China from A guide to the mammals of China<sup>16</sup> and Catalogue of mammals in China (2021)<sup>24</sup>.

Two researchers independently reviewed the information of each species of rodent and shrew, and entered the data into a standardized sheet to establish a database. Discrepancies were resolved by discussion between the

two researchers and facilitated by a third senior researcher to reach consensus, and for those records where the time of discovery and location were missing in the articles, we contacted the corresponding authors for detailed information. The species lists were reviewed by specialists in small mammal groups<sup>20</sup>. Detailed descriptions for literature search and schematic procedures are provided in the Fig. 1.

Using the key words searching, a total of 58,021 papers were retrieved for screening, comprised of 15,959 papers from English database and 42,062 papers from Chinese database. For the second-step screening, the abstracts of all returned references were reviewed to exclude those reporting solely clinical case or laboratory data or reporting diagnostic development, without mentioning any specific rodent species. For the third step screening, we reviewed the full text of all the remaining 720 papers in detail, from which a total of 430 Chinese papers and 290 English papers were determined to be eligible for data extraction (Fig. 1). The earliest Chinese and English publications were published in 1958 and 1986, respectively, and an abrupt increase of literature with rodent and shrew records was seen after 1980 (Fig. 2). The key data were extracted from the obtained papers, books, field surveys and websites: (i) species name of rodents and shrews, (ii) the reported geographic location information at province, city and county levels, (iii) time of identification and reporting, (iv) source of papers, books, field surveys and websites. All the data were entered into an excel spreadsheet for downstream analysis. After the initial data entrance, a twice re-check by two persons was performed to correct errors and remove duplicates. Particularly, historical changes in taxonomy of rodents and shrews were considered and a standardized terminology was applied for the same species<sup>24</sup>. Where necessary, names of study sites that were historically used were updated to the current name. In total, 13,911 records were compiled, with several records within each province shown across the study years (Fig. 2).

**Geo-positioning.** Location information was manually extracted at the highest resolution for each record of rodents and shrews, and then was categorized into three administrative levels such as province, city, and county. For a species, we only remain one record at the county level if various geographic locations (e.g., township, village) were reported within a county, and then we deleted the duplications for those records in the same county from different studies. For publications that reported neither specific administrative region nor coordinates, e.g., a scenic area or a mountain, we searched the coordinates for investigation sites by using Baidu Map (<https://map.baidu.com/>) and recorded their coordinates, which were linked to the county map by using the ArcGIS 10.7 software (ESRI Inc., Redlands, CA, USA) so as to gain the location information at county level. China administrative boundary maps of province, city and county (2015) were collected from the Resource and Environment Data Cloud Platform, Chinese Academy of Sciences (<http://www.resdc.cn>). A total of 13,911 records were identified including 1,419, 1,087, and 11,405 records respectively at the province, city, and county level. A “Location level” field was used to accurately demonstrate the spatial resolution such as province, city or county for each record in our dataset. This classification allowed for separate extraction and usage at different spatial levels for various purposes. We used RStudio Version 1.4.1103 and ArcGIS 10.7 software to analyze and visualize the obtained geographic data statistically.

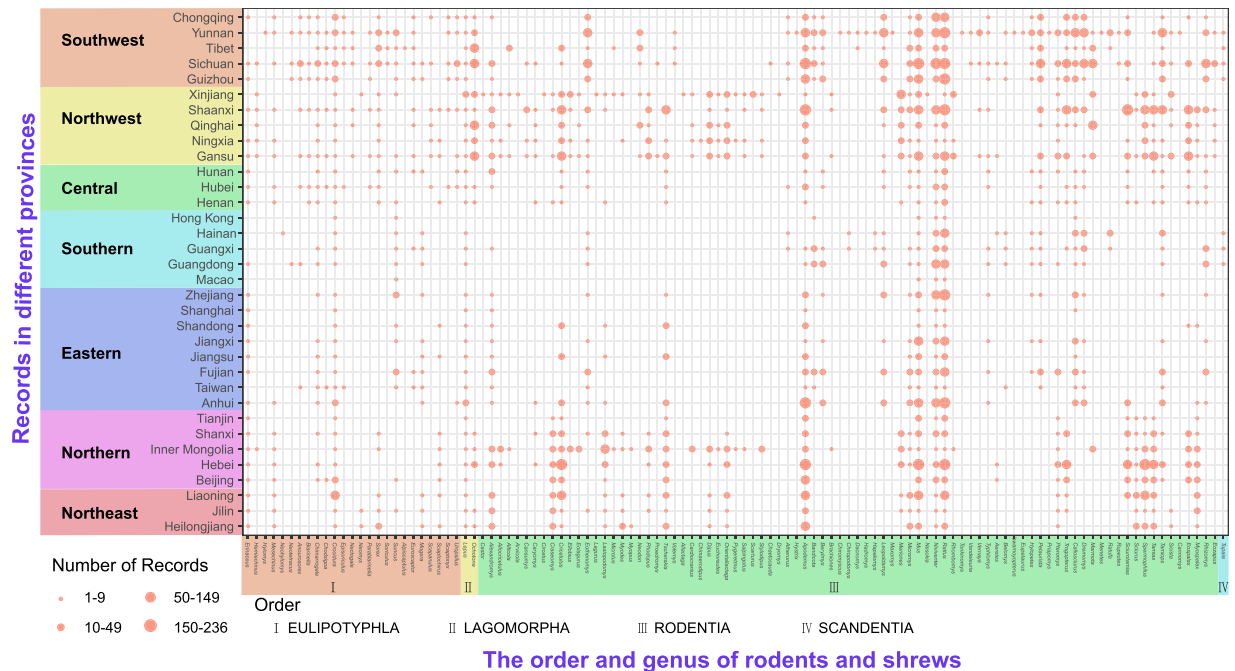
## Data Records

In this distribution and diversity dataset of rodents and shrews in China, as accessible from figshare<sup>25</sup>, each dataset row describes a distinct record (an occurrence of rodents and shrews in a specific location as described at a set time-point). The database contains the following fields:

1. **ID:** Unique identifier code of the database records.
2. **Species Authority:** The authority or author credited with the first formal use of the species scientific name.
3. **Order:** Identification of order of rodents and shrews.
4. **Family:** Identification of family of rodents and shrews.
5. **Genus:** Identification of genus of rodents and shrews.
6. **Species:** Identification of species of rodent and shrew.
7. **English Name:** English names of species of rodent and shrew.
8. **Chinese Name:** Chinese names of species of rodent and shrew.
9. **Synonyms:** Different names of species of rodent and shrew.
10. **China's Red List of Biodiversity (2021):** The protection status assessment of species of rodents and shrews in the International Union for Conservation of Nature (IUCN) category.
11. **Assessment information:** The assessment information section includes the year assessed, assessors and reviewers of species of rodent and shrew.
12. **Assessment Criteria:** The assessment criteria is displayed for taxa assessed as threatened (i.e., taxa evaluated as Critically Endangered, Endangered or Vulnerable), and standards that are nearly met are displayed for Near Threatened taxa.
13. **Distribution Note:** The non-endemic and endemic of the species of rodent and shrew.
14. **Location level:** The geographic scale of location (1 = province level, 2 = city level, 3 = county level).
15. **Province:** Province level information of rodents and shrews reported location, including names of 34 provinces, autonomous regions and municipalities of China.
16. **City:** City level information of rodents and shrews reported location, including names of 304 prefectural-level cities or autonomous prefectures.
17. **County:** County level information of rodents and shrews reported location, including names of 1,660 counties.
18. **Province code:** Adopt 2015 China's administrative division province name code.
19. **City code:** Adopt 2015 China's administrative division city name code.







**Fig. 3** Number of genus records for different rodents and shrews at province level in the mainland of China.

### Technical Validation

Herein, this dataset contains 13,911 records extracted from 720 kinds of literature that were published between 1950 and 2021, as well as 2,188 additional effective records from *Fauna sinica, mammal, vol.6(2), cricetidae*<sup>15</sup>, *A guide to the mammals of China*<sup>16</sup>, *Distribution of mammal species in China*<sup>17</sup>, *Colored atlas of Chinese mammals*<sup>18</sup>, *Mammals of Tibet*<sup>19</sup>, *China's red list of biodiversity: vertebrates, volume I, mammal*<sup>20</sup>, *Atlas of epidemiology of natural focus diseases in China*<sup>21</sup>, *Fauna of Beijing*<sup>22</sup>, *Fauna of Guizhou*<sup>23</sup>, *Handbook of the mammals of the world. volume 8*<sup>26</sup>, *Mammal species of the world: a taxonomic and geographic reference. volume 3*<sup>27</sup>, etc.

A senior investigator supervised the data entry performed by three data extraction analysts, two members for the records entrance and double check by each other, one member for the validation, emphasizing removing duplicate data, standardizing the formatting and naming conventions across the data. Due to the considerable period of the literature search over 70 years, the classification and naming of rodents and shrews have changed, and we have taken “*Catalogue of mammals in China (2021)*<sup>24</sup> and *A complete checklist of mammal species and subspecies in China: a taxonomic and the geographic reference*<sup>28</sup> as the standard criterion for the naming conventions. Records were checked strictly to ensure accuracy and extraction criteria were met, following a previously described approach used in *Herrera et al.*<sup>29</sup> and *Meng et al.*<sup>30</sup>. Data were checked strictly to ensure that accuracy and extraction criteria were met, similar to the method used in *Zhang et al.*<sup>31</sup>.

The final georeferenced dataset contained 364 rodents and shrews in 4 orders, 18 families and 114 genera. We listed the catalogue, medical importance, and ecological characteristics of each kind of rodents and shrews in the Supplement Table 1, and data concerning medical importance and ecological characteristic were collected from “*A catalogue and geographical distribution of the vectors of China*<sup>32</sup> and *China's red list of biodiversity: vertebrates, volume I, mammal*<sup>20</sup>. Figure 3 shows the diversity and number of records of all the rodents and shrews at the province level. Meanwhile, China has been divided into seven geographical regions, for each record, even though a rodents and shrews may be reported to be present in different locations, only one point was marked. As a result, the 18 rodent and shrew families were identified, and their distributions were illustrated in Fig. 4(a–m). The gray planes represent provincial records, the area covered by slash represents city records, and the light blue outline represent county records. The geographical locations have been color coded to represent the order of EULIPOTYPHILA, LAGOMORPHA, RODENTIA and SCANDENTIA. At the same time, the distribution density of rodents and shrews occurrences across different Eco-geographical fauna, climatic zones and vegetation zones are displayed and summarised in Fig. 4(n–p).

### Usage Notes

We provided a comprehensive and most updated comprehensive database for the geographic distribution and diversity of rodents and shrews at multiple scales in the mainland of China. The database contains 13,911 records, recording the distribution of 364 rodents and shrews that spanned from 1950 to 2021. Each record of the rodents and shrews was paired with relevant geo-positioning, thus enabling a future potential distribution prediction of rodents and shrews and disease risk assessment analysis.



## References

1. Meerburg, B. G., Singleton, G. R. & Kijlstra, A. Rodent-borne diseases and their risks for public health. *Crit Rev Microbiol.* **35**, 221 (2009).
2. NCBI SequenceRead Rabiee, M. H., Mahmoudi, A., Siahsarvie, R., Kryštufek, B. & Mostafavi, E. Rodent-borne diseases and their public health importance in Iran. *Plos Negl Trop Dis.* **12**, e0006256 (2018).
3. Wu, Z. Q. *et al.* Decoding the RNA viromes in rodent lungs provides new insight into the origin and evolutionary patterns of rodent-borne pathogens in Mainland Southeast Asia. *Microbiome.* **9** (2021).
4. Zheng, Z. M., Jiang, Z. K. & Chen, A. G. *Rodentology* (Shanghai Jiaotong Univ. Press, 2008).
5. Steffoff, R. *The Rodent Order* (Marshall Cavendish Press, 2008).
6. Wood, A. E. A revised classification of the Rodents. *J Mammal.* **36**, 165 (1955).
7. Morand, S. Emerging diseases, livestock expansion and biodiversity loss are positively related at global scale. *Biol Conserv.* **248**, 108707 (2020).
8. Vaheri, A. *et al.* Uncovering the mysteries of hantavirus infections. *Nat Rev Microbiol.* **11**, 539–550 (2013).
9. Wu, Z. Q. *et al.* Comparative analysis of rodent and small mammal viromes to better understand the wildlife origin of emerging infectious diseases. *Microbiome.* **6**, 178 (2018).
10. Jiang, Z. G. *et al.* China's mammalian diversity. *Biodiversity Science.* **23**, 351–364 (2015).
11. Mallapaty, S. China takes centre stage in global biodiversity push. *Nature.* **578**, 345–346 (2020).
12. Liu, J. J. *et al.* Forest fragmentation in China and its effect on biodiversity. *Biol Rev Camb Philos Soc.* **94**, 1636–1657 (2019).
13. Fang, L. Q. *et al.* Travel-related infections in mainland China, 2014–16: an active surveillance study. *Lancet Public health.* **3**, e385–e394 (2018).
14. Wu, Y. *et al.* Epidemiology of imported infectious diseases, China, 2014–18. *J Travel Med.* **27**, taaa211 (2020).
15. Luo, Z. X. *Fauna Sinica, Mammal, Vol.6(2), Cricetidae* (Science Press, 2000).
16. Smith, A. & Xie, Y. *A Guide to the Mammals of China* (Hunan Education Press, 2009).
17. Zhang, R. Z. *Distribution of mammal species in China* (China Forestry Press, 1997).
18. Pan, Q. H., Wang, Y. X. & Yan, K. *Colored atlas of Chinese mammals* (China Forestry Press, 2007).
19. Comprehensive scientific investigation team of Qinghai Tibet Plateau, Chinese Academy of Sciences. *Mammals of Tibet* (Science Press, 1986).
20. Jiang, Z. G. *China's Red List of Biodiversity: Vertebrates, Volume I, Mammals* (Science Press, 2021).
21. Cao, W. C., Fang, L. Q. & Wang, J. L. *Atlas of Epidemiology of Natural Epidemic Diseases in China* (Science Press, 2019).
22. Chen, W. *Fauna of Beijing* (Beijing Publishing house, 2002).
23. Luo, R. *Fauna of Guizhou* (Guizhou Science and Technology Publishing House, 1993).
24. Wei, F. W. *et al.* Catalogue of mammals in China (2021). *Journal of mammals.* **5**, 41 (2021).
25. Chen, J. J. *et al.* A dataset of diversity and distribution of rodents and shrews in China (1950–2021) v2. *Figshare* <https://doi.org/10.6084/m9.figshare.19608132.v1> (2022).
26. Russell, A. M. & Don. E. W. *Handbook of the Mammals of the World.* **8** (Lynx Edicions, 2018).
27. Wilson, D.E. & Reeder, D.M. *Mammal Species of the World: A Taxonomic and Geographic Reference* 3rd edn (The Johns Hopkins Univ. Press, 2005).
28. Wang, Y. X. *A Complete Checklist of Mammal Species and Subspecies in China: A Taxonomic and The Geographic Reference* (China Forestry Press, 2003).
29. Herrera, G. *et al.* An interactive database of Leishmania species distribution in the Americas. *Sci Data.* **7**, 110 (2020).
30. Meng, F. F. *et al.* A dataset of distribution and diversity of blood-sucking mites in China. *Sci Data.* **8**, 204 (2021).
31. Zhang, G. S., Zheng, D., Tian, Y. Q. & Li, S. A dataset of distribution and diversity of ticks in China. *Sci Data.* **6**, 105 (2019).
32. Xu, B. H. *A catalogue and geographical distribution of the vectors of China* (Fujian Science and Technology Publishing House, 2021).

## Acknowledgements

This work was funded by grants from the National Key Research and Development Project (No.2019YFC1200604) and the National Natural Science Foundation of China (82041022). We gratefully acknowledge all authors from the submitting laboratories who contributed to sharing information to literature, books, and websites, as most of our analyses were made possible by the sharing of their work. We also thank Yanhe Wang for assistance with data collection.

## Author contributions

All the authors in the author list were involved in this paper's preparation and writing process, and all of them agreed to be responsible for all aspects of the research paper. W.L. and L.Q.F. designed this study; J.J.C., Q.X., F.F.M., T.W. and Z.W.L. performed experiments; J.J.C. analyzed the data and wrote the manuscript; T.W. and Z.W.L. contributed analysis methods and tools. J.J.C. and F.F.M. provided administrative guidance in this study. J.J.C., W.L. and L.Q.F. wrote the original draft. All of the authors reviewed the manuscript and approved the final version.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41597-022-01422-2>.

**Correspondence** and requests for materials should be addressed to L.-Q.F. or W.L.

**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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022