The Ecological Niche of bat Viruses Closely Related to SARS-CoV-2, as Inferred from Phylogeographic Analyses of Rhinolophus Species

Alexandre HASSANIN (✉ alexandre.hassanin@mnhn.fr)
 Sorbonne University

Vuong Tan Tu
 Vietnam Academy of Science and Technology

Manon Curaudeau
 Sorbonne University

Gabor Csorba
 Hungarian Natural History Museum

The ecological niche of bat viruses closely related to SARS-CoV-2, as inferred from 1

phylogeographic analyses of *Rhinolophus* species 2

3

Alexandre Hassanin[1°]*, Vuong Tan Tu[2°], Manon Curaudeau[3], and Gabor Csorba[4] 4

5

[1] Institut de Systématique, Évolution, Biodiversité (ISYEB), Sorbonne Université, MNHN, CNRS, 6

EPHE, UA, Paris; alexandre.hassanin@mnhn.fr 7

[2] Institute of Ecology and Biological Resources, Vietnam Academy of Science and Technology, 8

No. 18, Hoang Quoc Viet Road, Cau Giay District, Hanoi, Vietnam; tuvuongtan@gmail.com 9

[3] Institut de Systématique, Évolution, Biodiversité (ISYEB), Sorbonne Université, MNHN, CNRS, 10

EPHE, UA, Paris; manon.curaudeau1@mnhn.fr 11

[4] Department of Zoology, Hungarian Natural History Museum, Baross u. 13, Budapest H-1088, 12

Hungary; csorba.gabor@nhmus.hu 13

° co-first authors 14

* correspondence: alexandre.hassanin@mnhn.fr; Tel.: +33 1 40 79 56 93 15

16

Abstract

To date, viruses closely related to SARS-CoV-2 have been reported in four bat species: *Rhinolophus acuminatus*, *Rhinolophus affinis*, *Rhinolophus malayanus*, and *Rhinolophus shameli*. Here, we analysed 343 sequences of the mitochondrial cytochrome c oxidase subunit 1 gene (*CO1*) from georeferenced bats of the four *Rhinolophus* species identified as reservoirs of viruses closely related to SARS-CoV-2. Haplotype networks were constructed in order to investigate patterns of genetic diversity among bat populations of Southeast Asia and China. No strong geographic structure was found for the four *Rhinolophus* species, suggesting high dispersal capacity. The ecological niche of bat viruses closely related to SARS-CoV-2 was predicted using the four localities in which bat viruses were recently discovered and the localities where bats showed the same *CO1* haplotypes than virus-positive bats. The ecological niche of bat viruses related to SARS-CoV was deduced from the localities where bat viruses were previously detected. The results show that the ecological niche of bat viruses related to SARS-CoV2 includes several regions of mainland Southeast Asia whereas that of bat viruses related to SARS-CoV is mainly restricted to China. In agreement with these results, human populations in Laos, Vietnam, Cambodia, and Thailand appear to be much less affected by the Covid-19 pandemic than other countries of Southeast Asia. In the climatic transitional zone between the two ecological niches (southern Yunnan, northern Laos, northern Vietnam), genomic recombination between highly divergent viruses is more likely to occur. Since recombinant viruses can threaten the benefit of vaccination campaigns, these regions should be under surveillance.

Keywords: coronavirus; horseshoe bats; reservoir hosts; Indochina; China; immunity.

Introduction

The Severe Acute Respiratory Syndrome coronavirus 2 (SARS-CoV-2) emerged in December 41
2019 in Wuhan (China) [1]. After 16 months, the coronavirus disease 2019 (Covid-19) pandemic 42
has affected more than 137 million of people around the world, claiming over 2.95 million lives 43
[2]. The origin of SARS-CoV-2 remains enigmatic and many hypotheses have been advanced to 44
explain the first animal-to-human transmission [3]. 45

Within the family Coronaviridae, the subgenus *Sarbecovirus* includes two human viruses, SARS- 46
CoV-2 and SARS-CoV (which was responsible for the SARS epidemic in 2002-2004) [4]. The 47
genomes of these two viruses share only 80% of nucleotide identity, and whole-genome 48
phylogenies have shown that they belong to two divergent lineages [1, 5-7], which we refer to 49
herein as SARS-CoV related coronaviruses (SCoVrCs) and SARS-CoV-2 related coronaviruses 50
(SCoV2rCs). Most SCoVrCs were discovered in bats collected in China after the SARS epidemic, 51
and the great majority were found in horseshoe bat species of the genus *Rhinolophus* (family 52
Rhinolophidae), suggesting that this taxon is the natural reservoir host of sarbecoviruses [8]. More 53
recently, several viruses showing between 96.2 and 91.8 % of genome identity with SARS-CoV-2 54
were identified from fecal samples of four horseshoe bat species: the RaTG13 virus was isolated 55
from a *R. affinis* bat collected in Yunnan in 2013 [1]; the RmYN02 virus was found in a *R.* 56
malayanus bat sampled in Yunnan in 2019 [5]; two variants of the same virus RshSTT200 (other 57
variant: RshSTT182) were detected in two *R. shameli* bats caught in northern Cambodia in 2010 58
[6]; and five variants of the same virus RacCS203 (other variants: RacCS224, RacCS253, 59
RacCS264, and RacCS271) were sequenced from five *R. acuminatus* bats collected in eastern 60
Thailand in 2020 [7]. The bat species *R. acuminatus* and *R. shameli* are endemic to Southeast Asia 61
whereas the two other bat species, *R. affinis* and *R. malayanus*, are distributed in both Southeast 62
Asia and China (Fig. 1), suggesting that Southeast Asia is the main region where bat SCoV2rCs 63
diversified. In addition, these recent data confirmed that the genus *Rhinolophus* is the natural 64
reservoir host of all sarbecoviruses [3,8]. Note that this hypothesis was already corroborated by the 65

discovery of two divergent sarbecovirus genomes in horseshoe bat species collected out of Asia: 66

one in *Rhinolophus blasii* from Bulgaria (BM48-31) [9] and another in an unidentified *Rhinolophus* 67

species from Kenya (BtKY72) [10]. 68
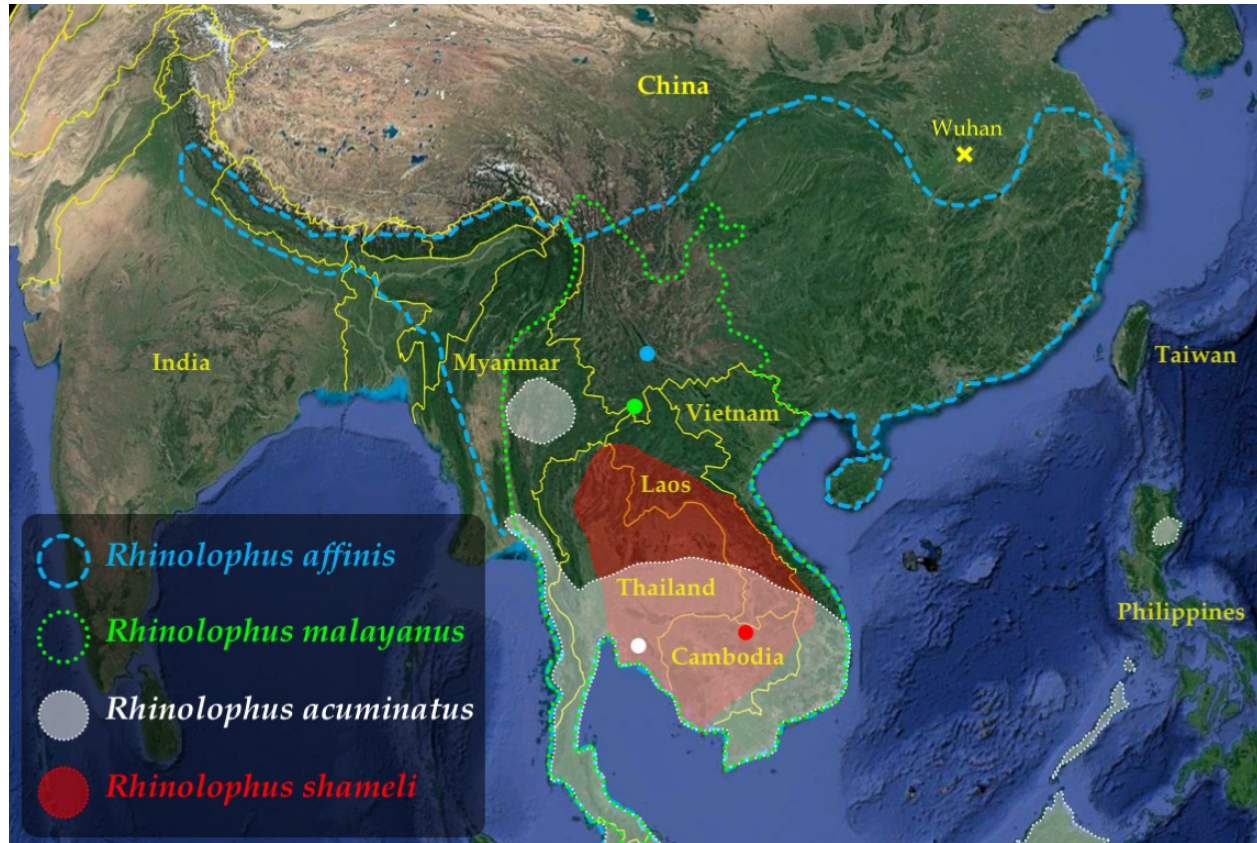
69

Figure 1. Geographic distribution of the four *Rhinolophus* species found positive for viruses 70
closely related to SARS-CoV-2 in southern China and Southeast Asia. Map from Google Earth 71
US Dept of State Geographer © 2020 Google – Image Landsat / Copernicus - Data SIO, NOAA, 72
U.S. Navy, NGA, GEBCO. For each of the four *Rhinolophus* species, the geographic distribution 73
was extracted from the IUCN website [11]. The coloured dots show the four geographic locations 74
where bats found positive for SCoV2rCs were collected. 75

76

Since SCoV2rCs have been circulating in horseshoe bats for many decades [12], it is important to 77

study population genetic structure of bats found positive for these sarbecoviruses in order to 78

evidence their dispersal capacity in China and Southeast Asia. In this report, the phylogeography 79

of the four species *R. acuminatus*, *R. affinis*, *R. malayanus* and *R. shameli* was therefore examined 80

using the mitochondrial cytochrome c oxidase subunit 1 gene (CO1) from 343 georeferenced bats 81

collected in 62 localities of Southeast Asia and China (Fig. 2). For each of the four species, 82

haplotype networks were constructed to investigate geographic patterns of genetic diversity among 83

bat populations. The results of these analyses were used to select specific location coordinates to 84

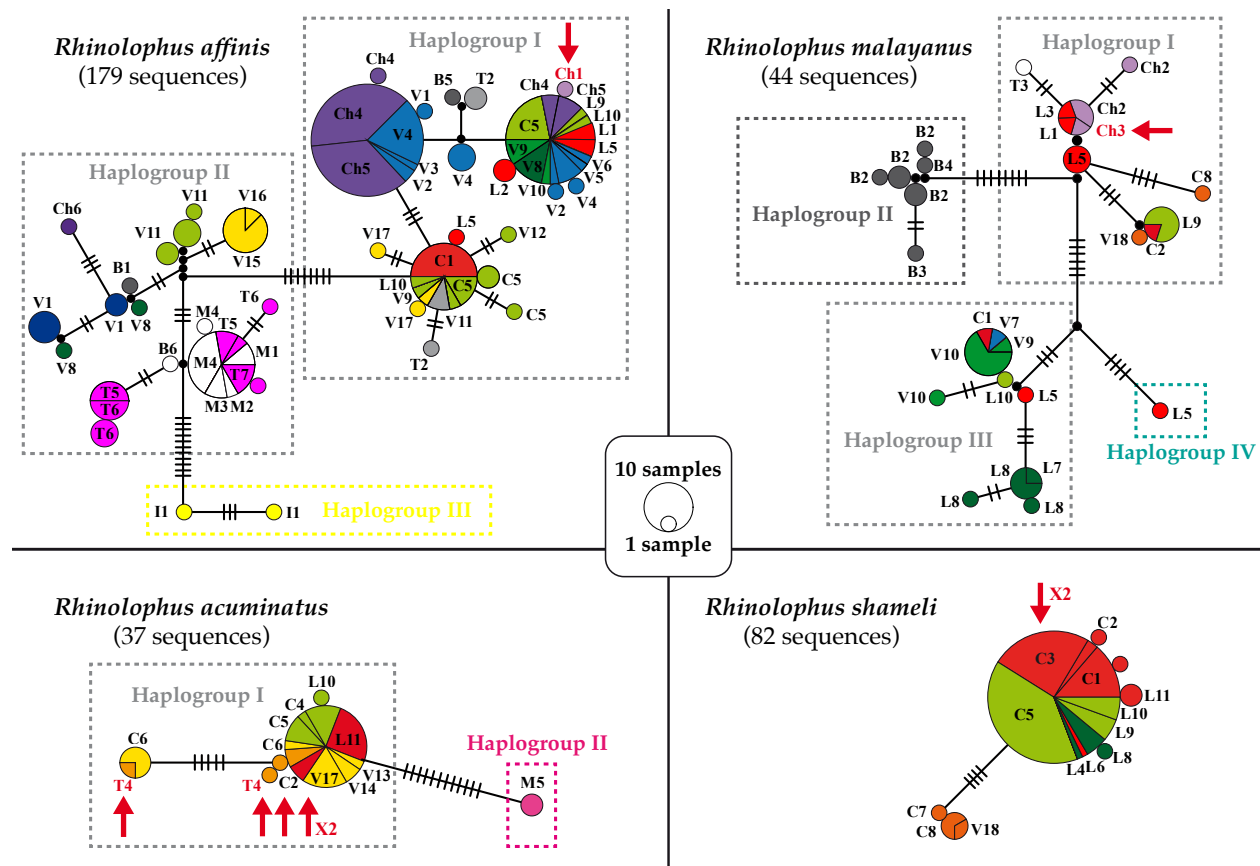predict the ecological niche of bat SCoV2rCs. 85
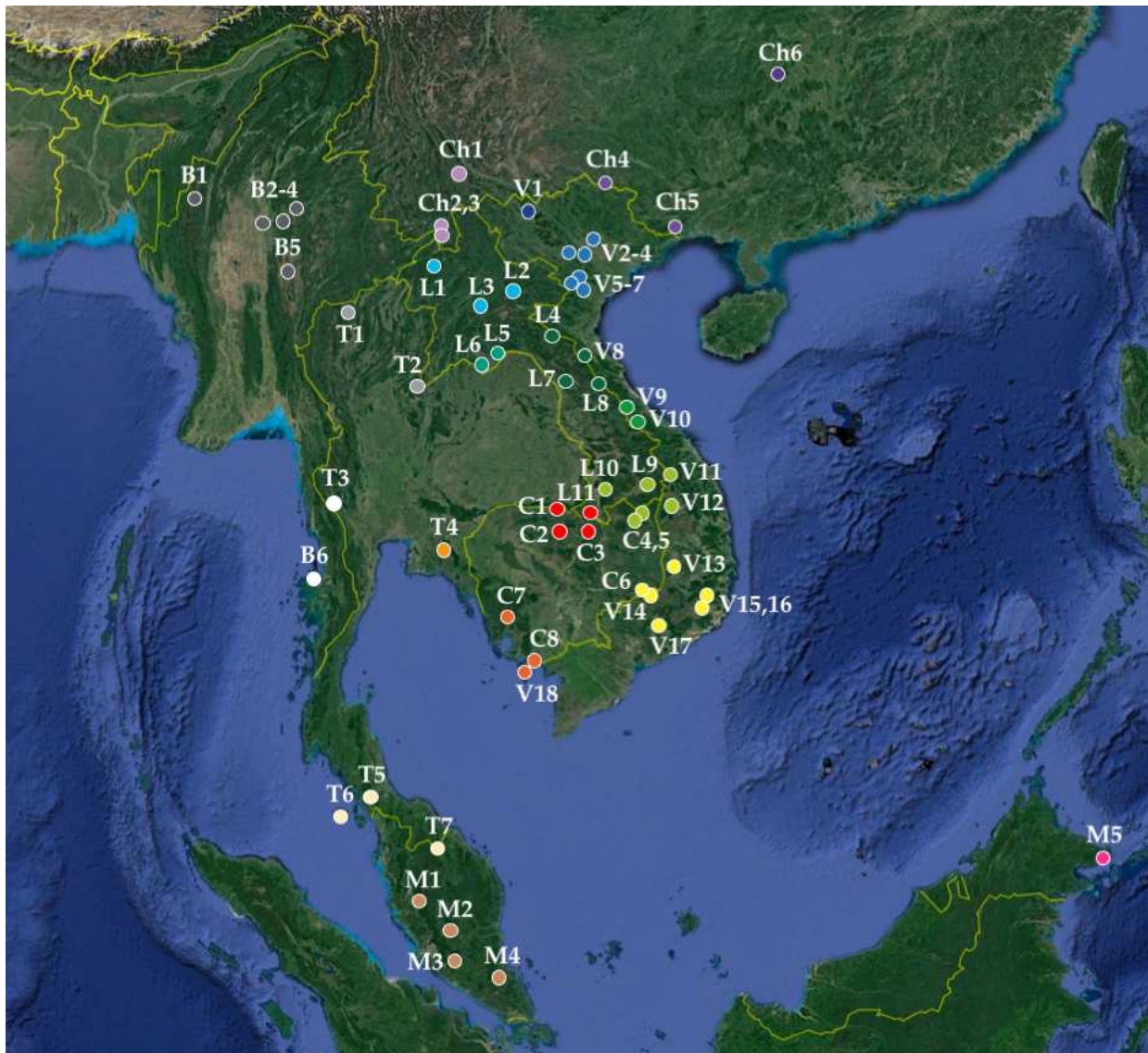
86

Figure 2. Geographic localities analysed in this study. The codes used for the countries are the 87
following: B (Myanmar), C (Cambodia), Ch (China), I (Indonesia), L (Laos), M (Malaysia), T 88
(Thailand), and V (Vietnam). Different regions were highlighted using colours. Map from Google 89
Earth US Dept of State Geographer © 2020 Google – Image Landsat / Copernicus - Data SIO, 90
NOAA, U.S. Navy, NGA, GEBCO. The locality names and GPS coordinates are provided in online 91
supplementary table S1. 92

93

Results and Discussion 94

5

Genetic analyses of *Rhinolophus* species identified as reservoirs of viruses closely related to SARS-CoV-2

Until now, SCoV2rCs have been found in four bat species of the genus *Rhinolophus*: *R. acuminatus*, *R. affinis*, *R. malayanus*, and *R. shameli*. The haplotype networks constructed using *CO1* sequences of these four species are shown in Fig. 3. A star-like genetic pattern, characterized by one dominant haplotype and several satellite haplotypes was found for the two bat species endemic to Southeast Asia, i.e. *R. acuminatus* and *R. shameli*.

Figure 3. Haplotype networks based on *CO1* sequences of the four *Rhinolophus* species found positive for viruses closely related to SARS-CoV-2 (SCoV2rCs). The networks were constructed with the median joining method available in PopART 1.5 [13]. The codes used for the countries are the following: B (Myanmar), C (Cambodia), Ch (China), I (Indonesia), L (Laos), M (Malaysia), T (Thailand), and V (Vietnam). Colours indicate the geographic origin of haplotypes according to Fig. 2 (see online supplementary table S1). The circles indicate haplotypes separated by at least one mutation. The black lines on the branches show the number of mutations ≥ 2. Black circles represent missing haplotypes. Circle size is proportional to the number of haplotypes. Haplogroups separated by more than seven mutations (pairwise nucleotide distances > 1%) are highlighted by dotted lines. The red arrows show the positions of the nine bats found positive for SCoV2rCs.

In the network of *R. acuminatus*, the most common haplotype (named Rac1 in online supplementary table S1) was found in northern Cambodia, southern Laos, eastern Thailand and southern Vietnam, indicating recent gene flow among these populations. Since a virus related to

6

SARS-CoV-2 (91.8% of genome identity), named RacCS203, was detected in five *R. acuminatus* 116
bats caught in eastern Thailand in June 2020 [6], the genetic pattern obtained for this species 117
suggests that viruses closely related to RacCS203 may have circulated in most southern regions of 118
mainland Southeast Asia. In contrast, *R. acuminatus* bats collected in Borneo (M5) showed a 119
divergent haplotype (separated by 12 mutations; haplogroup II), suggesting that the South China 120
Sea between mainland Southeast Asia and Borneo constitutes a barrier to gene flow. Isolated 121
populations of *R. acuminatus* described in northern Myanmar, Indonesia (Java and Sumatra) and 122
the Philippines [14] should be further studied. 123

The network of *R. shameli* shows a typical star-like pattern, the most common haplotype (named 124
Rsh1 in online supplementary table S1) being detected in northern Cambodia and Laos. Since a 125
virus related to SARS-CoV-2 (93.1% of genome identity), named RshSTT200, was recently 126
discovered in two *R. shameli* bats collected in northern Cambodia in December 2010 [7], the 127
genetic pattern obtained for this species suggests that viruses closely related to RshSTT200 may 128
have circulated, at least in the zone between northern Cambodia and central Laos. The bats sampled 129
south to the Tonle Sap lake (n = 4; southern Cambodia and Vietnamese island of Phu Quoc) were 130
found to be genetically isolated from northern populations (four mutations). However, further 131
sampling in the south is required to confirm this result, as it may reveal *CO1* sequences identical 132
to the haplotypes detected in the north. 133

For the two species distributed in both China and Southeast Asia, i.e. *R. affinis* and *R. malayanus*, 134
the genetic patterns are more complex with different haplogroups showing more than 1% of 135
nucleotide divergence. In the network of *R. affinis*, there are three major haplogroups (named I, II 136
and III in Fig. 3) separated by a minimum of seven mutations. The results are therefore in agreement 137
with those previously published using *CO1* and D-loop mitochondrial sequences [15]. The *CO1* 138
haplotypes detected in the localities sampled in southern China (ch1, ch4, ch5) are distantly related 139
to the single haplotype available for central China (ch6), but they are also found in Laos, northern 140
and central Vietnam, northern Thailand and northeastern Myanmar. This result suggests recent 141

gene flow between populations from southern Yunnan and those from northern mainland Southeast 142
Asia. Since a virus related to SARS-CoV-2 (96.2% of genome identity), named RaTG13, was 143
detected in one *R. affinis* bat captured in southern Yunnan in 2013 [1], the genetic pattern obtained 144
for this species suggests that viruses closely related to RaTG13 may have circulated in the zone 145
comprising southern Yunnan and northern mainland Southeast Asia. 146

In the network of *R. malayanus,* there are four major haplogroups (named I, II, III and IV in Fig. 147
3) separated by a minimum of seven mutations. The *CO1* haplotypes detected in the localities 148
sampled in southern China (ch2 and ch3) were also found in northern Laos (L1 and L3), suggesting 149
recent gene flow between populations from these two countries. Since a virus related to SARS-150
CoV-2 (93.7% of genome identity), named RmYN02, was recently isolated from one *R. malayanus* 151
bat collected in southern Yunnan in June 2019 [5], the genetic pattern obtained for this species 152
suggests that viruses closely related to RmYN02 may have circulated, at least between southern 153
Yunnan and northern Laos. In contrast, the bats sampled in Myanmar were found to be genetically 154
isolated from other geographic populations (haplogroup II in Fig. 3). 155

156

Two different ecological niches for bat viruses related to either SARS-CoV or SARS-CoV-2 157
Sarbecoviruses have been already detected in a wide diversity of horseshoe bat species collected 158
in Asia, Africa and Europe [5-10]. Therefore, there is no doubt that *Rhinolophus* species constitute 159
the natural reservoir host of all sarbecoviruses [3,8]. The genus *Rhinolophus* currently includes 160
between 92 [11] and 109 [16] insectivorous species that inhabit temperate and tropical regions of 161
the Old World, with a higher biodiversity in Asia (63-68 out of the 92-109 described species) than 162
in Africa (34-38 species), Europe (5 species) and Oceania (5 species). Although some *Rhinolophus* 163
species are solitary, most of them are gregarious and live in large colonies or small groups generally 164
in caves and hollow trees, but also in burrows, tunnels, abandonned mines, and old buildings 165
[11,16]. However, they prefer large caves with total darkness, where temperatures are stable and 166
less affected by diurnal and seasonal climatic variations. Importantly, all *Rhinolophus* species in 167

which sarbecoviruses were detected in previous studies [1, 5-9; 17] are cave species that form small   168
groups or colonies (up to several hundreds) [11,18,19].   169
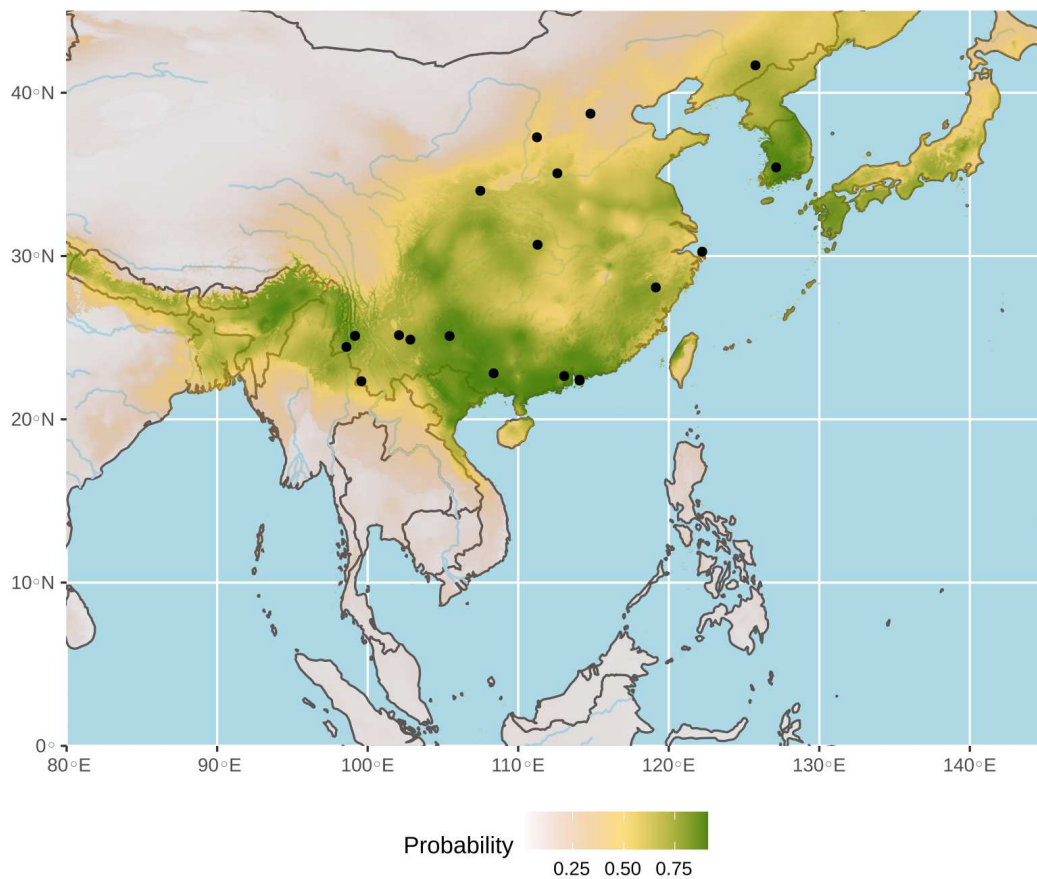
Figure 4. Ecological niche of bat viruses related to SARS-CoV (SCoVrCs). The geographic   171
distribution of suitable environments was predicted using the Maxent algorithm in ENMTools (see   172
Methods for details). AUC = 0.81. Black circles indicate localities used to build the distribution   173
model (see geographic coordinates in online supplementary table S2).   174

175

In China, many SCoVrCs were previously detected in several horseshoe bat species, including   176

Rhinolophus sinicus, *Rhinolophus ferrumequinum* (currently *R. nippon*) [16], *Rhinolophus*   177

macrotis (currently *R. episcopus*) [16], *Rhinolophus pearsoni*, and *Rhinolophus pusillus*, and it has   178

been shown that they circulate not only among conspecific bats from the same colony, but also   179

between bat species inhabiting the same caves [17,20,21]. The ecological niche predicted for bat   180

SCoVrCs using a data set of 19 points (see online supplementary table S2) is shown in Fig. 4. The   181

AUC was 0.81. The value was > 95% CI null-model's AUCs (0.68), indicating that the model   182

performs significantly better than a random model (see online supplementary Fig. S1). The highest   183

probabilities of occurrence (highlighted in green in Fig. 4) were found in Nepal, Bhutan,   184

Bangladesh, northeastern India, northern Myanmar, northern Vietnam, most regions of China south 185

of the Yellow River, Taiwan, North and South Korea, and southern Japan. 186

187

In Southeast Asia and southern China, SCoV2rCs have currently been found in four *Rhinolophus* 188

species (*R. acuminatus*, *R. affinis*, *R. malayanus* and R. *shameli*) [1,6-8], but the greatest diversity 189

of horseshoe bat species in mainland Southeast Asia (between 28 and 36 species) [11,16] suggests 190

that many sarbecoviruses will be discovered soon. Despite the limited data currently available on 191

SCoV2rCs, several arguments support that bat intraspecific and interspecific transmissions also 192

occur with SCoV2rCs. Firslty, recent genomic studies have revealed that SCoV2rCs circulate and 193

evolve among horseshoe bats of the same colony, as five very similar genomes (nucleotide 194

distances between 0.03% and 0.10%) were sequenced from five *R. acuminatus* bats collected from 195

the same colony in eastern Thailand [6], and as two genomes differing at only three nucleotide 196

positions (distance = 0.01%) were sequenced from two *R. shameli* bats collected at the same cave 197

entrance on the same night [7]. Secondly, the discovery of four viruses closely related to SARS- 198

CoV-2 (between 96.2 and 91.8 % of genome identity) in four different species of *Rhinolophus* is a 199

strong evidence that interspecific transmission occurred several times in the past. As detailed in 200

online supplementary Table S1, these species were collected together in several localities of 201

Cambodia (three species in C1, C2, and C5; two species in C8), Laos (four species in L10; three 202

species in L9; two species in L1, L5, L8, L11), and Vietnam (two speciess in V10, V9, V17, V18). 203

These data corroborate previous studies suggesting that sarbecoviruses can be transmitted, at least 204

occasionally, between *Rhinolophus* species sharing the same caves. 205
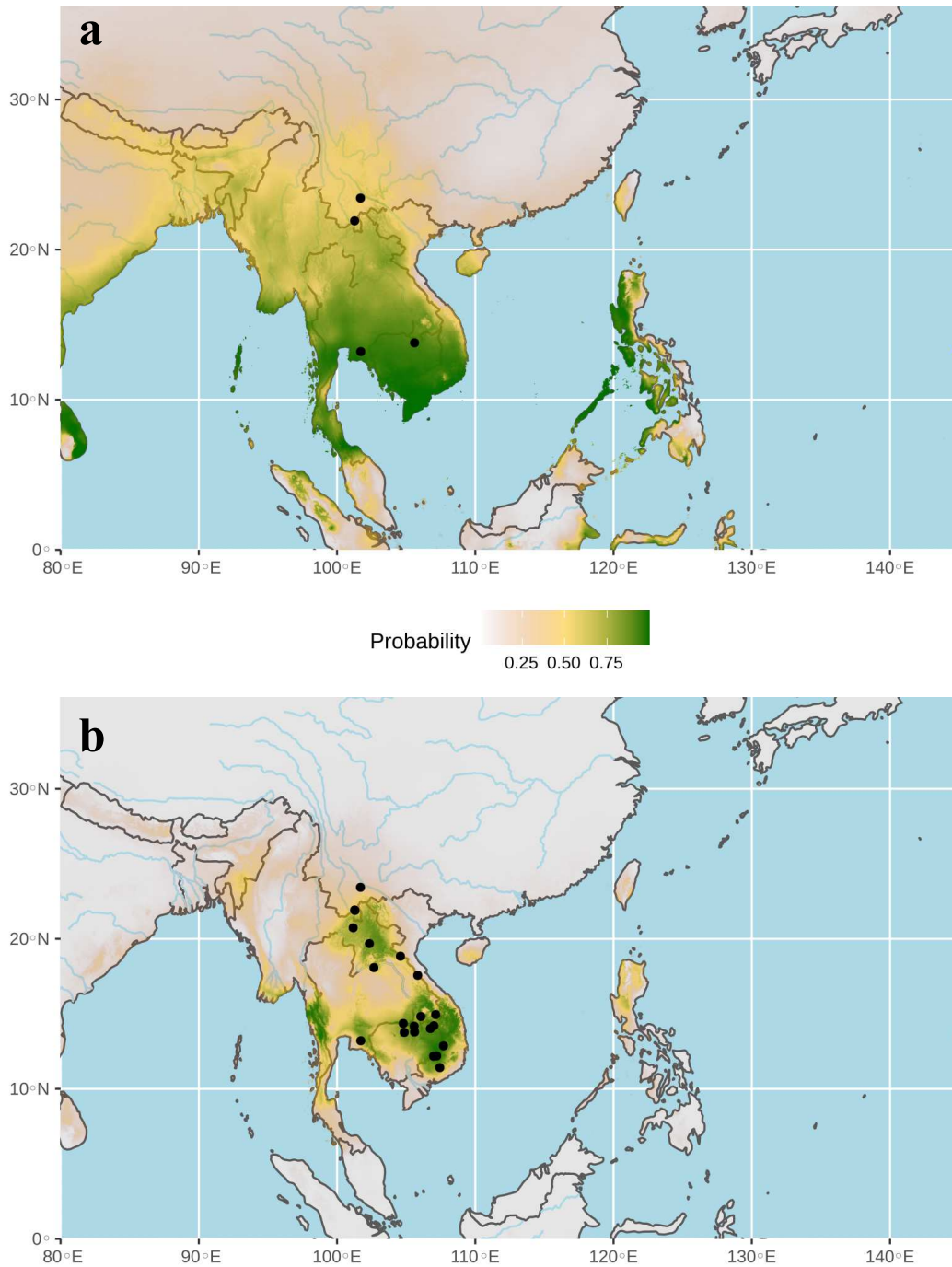
10

Figure 5. Ecological niches of bat viruses closely related to SARS-CoV-2 (SCoV2rCs) predicted using 4 points (a) (AUC = 0.58) and 21 points (b) (AUC = 0.96). The geographic distributions of suitable environments were predicted using the Maxent algorithm in ENMTools (see Methods for details). Black circles indicate localities used to build the distribution model (see geographic coordinates in online supplementary table S1).

The ecological niche of bat SCoV2rCs was firstly predicted using the four localities where bat viruses were previously detected [1,6-8] (Fig. 5a). The highest probabilities of occurrence (highlighted in green in Fig. 5a) were found in Southeast Asia rather than in China. However, the

11

AUC was only 0.58, and the value was < 95% CI null-model's AUCs (0.74), indicating that the model was not statically supported at a significance level of 0.05 (see on line supplementary Fig. S2). As expected, these preliminary results confirmed that more than four records are needed to increase the accuracy of the distribution model [22]. For that reason, we used a genetic approach to determine geographic localities where bat SCoV2rCs are more likely to be found. The *CO1* sequences of the nine bats in which a SCoV2rC was detected are shown by red arrows in Fig. 3. For *R. affinis*, the *CO1* haplotype sequenced for the bat found positive for a SCoV2rC in southern Yunnan (site named Ch1 in Fig. 2 and 3) was not found in other sampled localities. For the three other bat species found positive for SCoV2rCs, identical *CO1* sequences were detected in bats from 17 other geographic localities (see online supplementary table S1). For *R. acuminatus*, the four *CO1* haplotypes sequenced for the bats found positive for SCoV2rCs in eastern Thailand (site named T4 in Fig. 2 and 3) were also found in four localities in Cambodia (C2, C4, C5, and C6), two localities in southern Laos (L10 and L11), and three localities in southern Vietnam (V13, V14, and V17). The results indicate high connectivity among *R. acuminatus* populations from eastern Thailand, Cambodia, southern Laos and southern Vietnam. For *R. malayanus*, the *CO1* haplotype sequenced for the bat found positive for a SCoV2rC in southern Yunnan (site named Ch3 in Fig. 2 and 3) was also found in another locality in southern Yunnan (Ch2) and two localities in northern Laos (L1 and L3). The results indicate high connectivity among *R. malayanus* populations from southern China and northern Laos. For *R. shameli*, the single *CO1* haplotype sequenced for the two bats found positive for SCoV2rCs in northern Cambodia (site named C3 in Fig. 2 and 3) was also found in three other localities in Cambodia (C1, C2, and C5) and five localities in Laos (L4, L6, L8, L9, and L10). The results indicate high connectivity among *R. shameli* populations from Cambodia and Laos. Based on these genetic data, the ecological niche of bat SCoV2rCs was predicted using 21 records corresponding to the four localities where bat viruses were previously detected [1,6-8] and the 17 localities where bats showed the same *CO1* haplotype than virus-positive bats (data set B: 21 points; see online supplemntary table S1 for details). The AUC was

0.96. The value was > 95% CI null-model's AUCs (0.81), indicating that the model performs significantly better than a random model (see online supplementary Fig. S3). The areas showing the highest probabilities of occurrence (highlighted in green in Fig. 5b) include southern Yunnan, northern and southern Laos, northwestern and southern Vietnam, northeastern and southwestern Cambodia, eastern, northern and western Thailand, and southern Myanmar.

Our results show that bat SCoVrCs and SCoV2rCs have different evological niches: that of SCoVrCs covers mainly China and several adjacent countries and extends to latitudes between 18° and 43°N, whereas that of SCoV2rCs covers northern mainland Southeast Asia and extends to latitudes between 10° and 24°N. Most *Rhinolophus* species involved in the ecological niche of SCoVrCs have to hibernate in winter when insect populations become significantly less abundant. This is not the case for most *Rhinolophus* species involved in the ecological niche of SCoVrC2s. Since this ecological difference may be crucial for the dynamics of viral transmission among bat populations, it needs to be further studied through comparative field surveys in different regions of China and Southeast Asia.

The ecological niches of SCoVrCs and SCoV2rCs slightly overlap in the zone including southern Yunnan, northern Laos, and northern Vietnam (Fig. 4 and Fig. 5b). This zone corresponds to the northern edge of tropical monsoon climate [23]. Highly divergent sarbecoviruses of the two main lineages SCoVrCs and SCoV2rCs are expected to be found in sympatry in this area. This is confirmed by the discovery of both SCoVrCs and SCoV2rCs in horseshoe bats collected in southern Yunnan [1,6,21]. Collectively, these data suggest that genomic recombination between viruses of the two divergent lineages are more likely to occur in bats roosting, at least seasonally, in the caves of these regions. Since highly recombinant viruses can threaten the benefit of vaccination campaigns, southern Yunnan, northern Laos, and northern Vietnam should be the targets of closer surveillance.
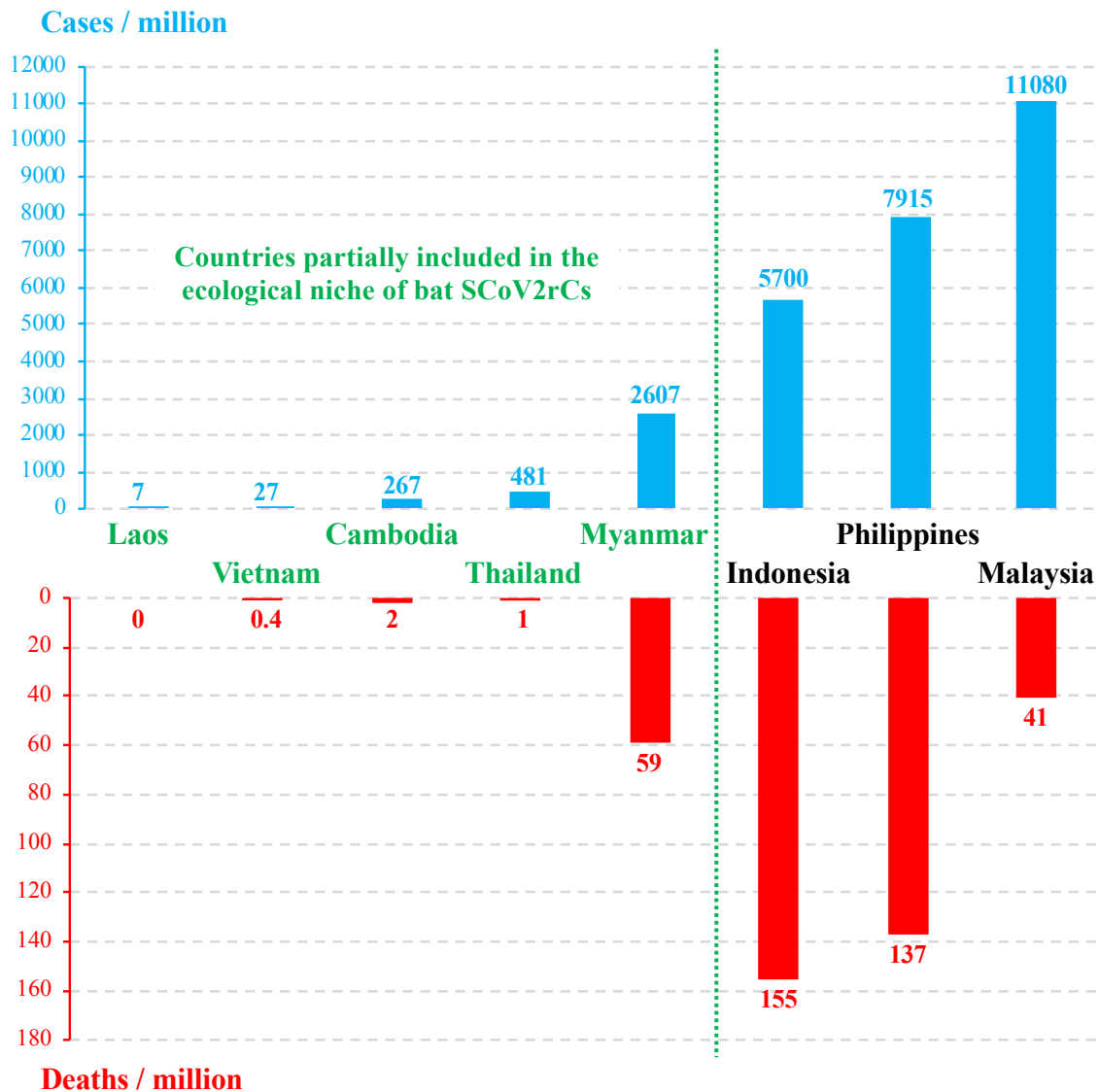
13

Figure 6. Number of Covid-19 patients per million inhabitants (in blue) and deaths per 270
million inhabitants (in red) for the different countries of Southeast Asia. Data extracted from 271
the Worldometers website [2] on April 12, 2021. 272

273

Chinese researchers have actively sought sarbecoviruses in all Chinese provinces after the 2002- 274

2004 SARS outbreak. They found many bat SCoVrCs [16,20,21] but only two SCoV2rCs [1,5] 275

and both of them were discovered in southern Yunnan, the Chinese province bordering Southeast 276

Asia. The ecological niches predicted herein for bat sarbecoviruses suggest that SCoVrCs are 277

dominant in China (Fig. 4) while SCoV2rCs are present mostly in Southeast Asia (Fig. 5). This 278

means that viruses similar to SARS-CoV-2 have been circulating for several decades throughout 279

Southeast Asia, and that different species of bats have exchanged these viruses in the caves they inhabit.

The data available on human cases and deaths caused by the Covid-19 pandemic [2] indirectly support the hypothesis that the cradle of diversification of bat SCoV2rCs is mainland Southeast Asia, and in particular the areas highlighted in green in Fig. 5b. Indeed, human populations in Cambodia, Laos, Thailand, and Vietnam appear to be much less affected by the Covid-19 pandemic than other countries of the region, such as Indonesia, Malaysia, Myanmar, and the Philippines (Fig. 6). This suggests that some human populations of Cambodia, Laos, Thailand, and Vietnam, in particular rural populations living in contact with wild animals for several generations, have a better immunity against SCoV2rCs because they have been regularly contaminated by bats and/or infected secondary hosts such as pangolins.

Pangolins contaminated by bats in Southeast Asia

Apart from bats, the Sunda pangolin (*Manis javanica*) and Chinese pangolin (*Manis pentadactyla*) are the only wild animals in which viruses related to SARS-CoV-2 have been found so far. However, these discoveries were made in a rather special context, that of pangolin trafficking. Several sick pangolins were seized by Chinese customs in Yunnan province in 2017 (unpublished data), in Guangxi province in 2017-2018 [24] and in Guangdong province in 2019 [25]. Even if the viruses sequenced in pangolins are not that close to SARS-CoV-2 (one was 85% identical and the other 90%), they indicate that at least two sarbecoviruses could have been imported into China well before the emergence of Covid-19 epidemic. Indeed, it has been shown that Sunda pangolins collected from different Southeast Asian regions have contaminated each other while in captivity on Chinese territory [3]. The question remained on how the Sunda pangolins became infected initially. Could it have been in their natural Southeast Asian environment, before being captured? The discovery of two new viruses close to SARS-CoV-2 in bats from Cambodia and Thailand [7,8] supports this hypothesis, as *Rhinolophus* bats and pangolins can meet, at least occasionally, in

15

forests of Southeast Asia, possibly in caves, tree hollows or burrows. Further substantiating this 306

hypothesis, the geographic distribution of *Manis javanica* [26] overlaps the ecological niche here 307

predicted for bat SCoV2rCs (Fig. 5), and SARS-CoV-2 neutralizing antibodies have been recently 308

detected in a Sunda pangolin collected in eastern Thailand [6]. Collectively, these data strengthen 309

the hypothesis that pangolin trafficking is responsible for multiple exports of viruses related to 310

SARS-CoV-2 to China [3]. 311

Methods 312

DNA extraction and sequencing 313

One hundred and forty-four tissue samples of morphologically identified bats of *R. acuminatus* (n 314

= 10), *R. affinis* (n = 57), *R. malayanus* (n = 14), and *R. shameli* (n = 63) were specially analysed 315

for this study. In the field, bats were captured with mist nets and harp-traps and handled according 316

to guidelines and recommendations of the American Society of Mammologists [27]. They were 317

measured, photographed and identified by the authors. Tissue samples were taken from the chest 318

muscles of voucher specimens or from the patagium (biopsy punches; 2 mm diameter) of released 319

bats. Samples were preserved in 95% ethanol. 320

Total DNA was extracted using QIAGEN DNeasy Tissue Kit (Qiagen, Germany) in accordance 321

with the manufacturer's instructions. The barcode fragment of the *CO1* gene (657 bp) was 322

amplified and sequenced using the primers UTyr and C1L705 [28]. PCR amplifications of the *CO1* 323

gene were performed as previously published [29]. PCR products were purified using ExoSAP Kit 324

(GE Healthcare, UK) and sequenced using the Sanger method on an ABI 3730 automatic sequencer 325

at the Centre National de Séquençage (Genoscope) in Evry (France). Haplotypes were assembled 326

with forward and reverse eletcropherograms using Sequencher 5.1 (Gene Codes Corporation, Ann 327

Arbor, MI, USA). No gaps and stop codons were found in the *CO1* sequences after translation into 328

amino-acids. Sequences generated for this study were deposited in the GenBank database 329

(accession numbers MW712891-MW713034) (see online supplementary table S1). 330

Analyses of CO1 sequences

Our sequences were aligned with 199 additional *CO1* sequences downloaded from GenBank. Note

that the *CO1* sequences of seven bats found positive for viruses closely related to SARS-CoV-2

[1,5,6] were assembled on Geneious® 10.2.2 (Biomatters Ltd., Auckland, New Zealand) by

mapping available SRA data to a *CO1* reference. Sequences were aligned using AliView 1.22 [30].

Our final *CO1* alignments contain 37 sequences for *R. acuminatus*, 44 sequences for *R. malayanus*,

82 sequences for *R. shameli*, and 180 sequences for *R. affinis*. These four alignments were analysed

in PopART 1.5 [13] to construct haplotype networks using the median joining method with equal

weights for all mutations. The 62 localities where bats were sampled are shown in the map of Fig.

2 and their geographic coordinates are detailed in online supplementary table S1.

Prediction of ecological niches

For bat SCoVrCs, the ecological niche was inferred using GPS data collected for viruses published

during the last two decades. The list of the 19 available geographic records is provided in online

supplementary table S2. For bat SCoV2rCs, the ecological niche was initially predicted using the

four geographic localities where viruses were previously detected [1,5-7]: two in Yunnan, one in

northern Cambodia, and one in eastern Thailand (data set A). However, the use of only four records

is questionable since Van Proosdij et al. [22] have estimated that a minimum of 13 records is

required to develop accurate distribution models for widespread taxa. For that reason, we used a

genetic approach to increase the number of geographic records. Since the detection of identical

CO1 sequences in different bat populations is indicative of recent dispersal events of females, we

also selected the 17 geographic records where bats showed the same *CO1* haplotypes than virus- 353

positive bats (data set B: 21 points; see online supplementary table S1). 354

For each of the three data sets (bat SCoVrCs; data sets A and B for bat SCoV2rCs), the 19 355

bioclimatic variables available in the WorldClim database [31] were studied for an area 356

corresponding to the minimum and maximum latitudes and longitudes of the selected points (19 357

points for bat SCoVrCs; 4 and 21 points, respectively for the SCoV2rCs data sets A and B) and the 358

caret R package [32] was used to determine the least correlated variables ($|r| < 0.7$) [33]. For bat 359

SCoVrCs, the following five predictor bioclimatic variables were retained: Bio3 (isothermality), 360

Bio4 (temperature seasonality), Bio5 (maximum temperature of the warmest month), Bio15 361

(precipitation seasonality), and Bio18 (precipitation of the warmest quarter). For data set A, the 362

following seven predictor bioclimatic variables were retained: Bio3, Bio7 (temperature annual 363

range), Bio10 (mean temperature of the warmest quarter), Bio13 (precipitation of the wettest 364

month), Bio14 (precipitation of driest month), Bio15, and Bio18. For data set B, the following 365

seven predictor bioclimatic variables were selected: Bio2 (mean diurnal range), Bio3, Bio7, Bio10, 366

Bio13, Bio15, Bio17 (precipitation of the driest quarter), and Bio18. Ecological niche modelling 367

was performed with the MaxEnt algorithm using ENMTools in R [34]. The MaxEnt approach was 368

chosen for its ability to work with presence-only data sets and to produce results with a low sample 369

size [35]. The area under the curve (AUC) of the receiver operating characteristic plot was used as 370

a first measure of model accuracy, a value of 0.5 indicating model accuracy not better than random, 371

and a value of 1 indicating perfect model fit [36,37]. To test for sampling bias, the distribution 372

model using all selected localities was tested against a null model developed by 1000 times drawing 373

an equal number of random points from the entire study area [37]. The position of the AUC value 374

was tested against the 95% confidence interval (CI) of the 1000 AUC values of the null-models. If   375

the AUC value is ≥ 95% CI null-model's AUCs, the model is considered performing significantly   376

better than a random model [37].   377

378

References   379

1. Zhou, P. *et al*. A pneumonia outbreak associated with a new coronavirus of probable bat origin.   380
    Nature **579**, 270-273; 10.1038/s41586-020-2012-7 (2020).   381

2. COVID-19 Coronavirus pandemic. https://www.worldometers.info/coronavirus/. downloaded   382
    on April 12, 2021.   383

3. Hassanin, A., Grandcolas, P. & Veron, G. Covid-19: natural or anthropic origin? *Mammalia*   384
    85, 1-7; 10.1515/mammalia-2020-0044 (2021).   385

4. Walker, P.J. *et al*. Changes to virus taxonomy and the statutes ratified by the international   386
    committee on taxonomy of viruses. *Arch Virol.* **165**, 2737-2748; 10.1007/s00705-020-04752-   387
    x (2020).   388

5. Zhou, H. *et al*. A novel bat coronavirus closely related to SARS-CoV-2 contains natural   389
    insertions at the S1/S2 cleavage site of the spike protein. *Curr. Biol.* **30**, 3896;   390
    10.1016/j.cub.2020.09.030 (2020).   391

6. Wacharapluesadee, S. *et al*. Evidence for SARS-CoV-2 related coronaviruses circulating in   392
    bats and pangolins in Southeast Asia. *Nat. Commun.* **12**, 972; 10.1038/s41467-021-21240-1   393
    (2021).   394

7. Hul, V. *et al*. A novel SARS-CoV-2 related coronavirus in bats from Cambodia.   395
    10.1101/2021.01.26.428212 (submitted).   396

8. Li, W. *et al*. Bats are natural reservoirs of SARS-like coronaviruses. *Science 310*, 676-679;   397
    10.1126/science.1118391 (2005).   398

9. Drexler, J.F. *et al*. Genomic characterization of severe acute respiratory syndrome-related 399
    coronavirus in European bats and classification of coronaviruses based on partial RNA- 400
    dependent RNA polymerase gene sequences. *J. Virol.* **84**, 11336-11349; 10.1128/JVI.00650- 401
    10 (2010). 402

10. Tao, Y. & Tong, S. Complete genome sequence of a severe acute respiratory syndrome-related 403
    coronavirus from Kenyan bats. *Microbiol. Resour. Announc.* **8**, e00548-19; 404
    10.1128/MRA.00548-19 (2019). 405

11. The IUCN Red List of Threatened Species. Version 2020-3. https://www.iucnredlist.org. 406
    Downloaded on 23 February 2021. 407

12. Boni, M.F. *et al*. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible 408
    for the COVID-19 pandemic. *Nat. Microbiol.* **5**, 1408–1417; 10.1038/s41564-020-0771-4 409
    (2020). 410

13. Leigh, J.W. & Bryant, D. PopART: Full-feature software for haplotype network construction. 411
    Methods Ecol. Evol. **6**, 1110-1116; 10.1111/2041-210X.12410 (2015). 412

14. Thong, V.D., Thanh, H.T., Soisook, P. & Csorba, G. *Rhinolophus acuminatus*. The IUCN Red 413
    List of Threatened Species e.T19520A21974227; 10.2305/IUCN.UK.2019- 414
    3.RLTS.T19520A21974227.en. (2019). 415

15. Ith, S. *et al*. Taxonomic implications of geographical variation in *Rhinolophus affinis* 416
    (Chiroptera: Rhinolophidae) in mainland Southeast Asia. *Zool. Stud.* **54**, e31; 10.1186/s40555- 417
    015-0109-8 (2015). 418

16. Burgin, C.J., Wilson, D.E., Mittermeier, R.A., Rylands, A.B., Lacher, T.E., Sechrest, W. 419
    Illustrated Checklist of the Mammals of the World. Vol. 2. Lynx Edicions, Barcelona (2020). 420

17. Fan, Y., Zhao, K., Shi, Z. L. & Zhou, P. Bat coronaviruses in China. *Viruses* **11**, 210; 421
    10.3390/v11030210 (2019). 422

18. Francis, C. Field Guide to the Mammals of South-east Asia. Bloomsbury, London, UK. (2019). 423

19. Smith, A.T. & Xie, Y. A Guide to the Mammals of China. Princeton University Press, Princeton, New Jersey (2008).

20. Han, Y. *et al*. Identification of diverse bat alphacoronaviruses and betacoronaviruses in China provides new insights into the evolution and origin of coronavirus-related diseases. *Front. Microbiol.* **10**, 1900; 10.3389/fmicb.2019.01900 (2019).

21. Hu, D. *et al*. Genomic characterization and infectivity of a novel SARS-like coronavirus in Chinese bats. *Emerg. Microbes Infect.*, **7**, 154; 10.1038/s41426-018-0155-5 (2018).

22. van Proosdij, A.S.J., Sosef, M.S.M., Wieringa, J.J. & Raes N. Minimum required number of specimen records to develop accurate species distribution models. *Ecography* **39**, 542-552; 10.5061/dryad.8sb8v (2016).

23. Song, Y., Achberger, C. & Linderholm, H.W. Rain-season trends in precipitation and their effect in different climate regions of China during 1961–2008. *Environ. Res. Lett*. **6**, 034025; 10.1088/1748-9326/6/3/034025 (2011).

24. Lam, T.T. *et al*. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature* **583**, 282-285; 10.1038/s41586-020-2169-0 (2020).

25. Liu, P., Chen, W. & Chen, J.P. Viral metagenomics revealed Sendai virus and coronavirus infection of Malayan pangolins (*Manis javanica*). *Viruses* **11**, 979; 10.3390/v11110979 (2019).

26. Challender, D. *et al*. *Manis javanica*. The IUCN Red List of Threatened Species 2019: e.T12763A123584856; 10.2305/IUCN.UK.2019-3.RLTS.T12763A123584856.en. (2019).

27. Sikes, R.S., Gannon W.L., and the animal care and use committee of the American Society of Mammalogists. Guidelines of the American Society of Mammalogists for the use of wild mammals in research. *J. Mammal.* **92**, 235–253; 10.1644/10-MAMM-F-355.1 (2011).

28. Hassanin, A. *et al*. Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. *C.R. biol*. **335**, 32–50; 10.1016/j.crvi.2011.11.002 (2012).

424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449

21

29. Tu, V.T. *et al.* Integrative taxonomy of the *Rhinolophus macrotis* complex (Chiroptera, Rhinolophidae) in Vietnam and nearby regions. *J. Zool. Syst. Evol. Res.* **55**, 177-198; 10.1111/jzs.12169 (2017).

30. Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large data sets. *Bioinformatics* **30**, 3276-3278; 10.1093/bioinformatics/btu531 (2014).

31. Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965-1978; 10.1002/joc.1276 (2005).

32. Kuhn, M. Building predictive models in R using the caret package. *J. Stat. Softw.* **28**, 1-26; 10.18637/jss.v028.i05 (2008).

33. Dormann, C.F. *et al.* Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **36**, 27-46; 10.1111/j.1600-0587.2012.07348.x (2013).

34. Warren, D.L. *et al.* ENMTools 1.0: an R package for comparative ecological biogeography. *Ecography* **44**, 504-511; 10.1111/ecog.05485 (2021).

35. Wisz, M.S. *et al.* Effects of sample size on the performance of species distribution models. *Divers. Distrib.* **14**, 763-773; 10.1111/j.1472-4642.2008.00482.x (2008).

36. Fielding, A.H. & Bell, J.F. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* **24**, 38-49; 10.1017/S0376892997000088 (1997).

37. Raes, N. & ter Steege, H. A null-model for significance testing of presence-only species distribution models. *Ecography* **30**, 727-736; 10.1111/j.2007.0906-7590.05041.x (2007).

Author Contributions 502

A.H. wrote the main manuscript text; V.T.T. and A.H. sequenced the CO1 barcodes; V.T.T. and 503

A.H. prepared table S1; A.H. prepared figures 1, 2, 3, and 6, and table S2; M.C. and A.H. prepared 504

figures 4, 5, and S1-S3; A.H, G.C., and V.T.T. collected bats in different localities of Southeast 505

Asia. All authors reviewed the manuscript. 506

507

Competing interests 508

The author(s) declare no competing interests. 509

510

Data Availability: DNA sequences generated for this study were deposited in the GenBank 511

database (accession numbers MW712891-MW713034). 512

513

Ethical statement: Ethical review and approval were not available for our study because the field 514

missions were carried out between 2004 and 2011, i.e., before the creation of the ethical committee 515

at the Muséum national d'Histoire naturelle. However, the field studies were carried out in 516
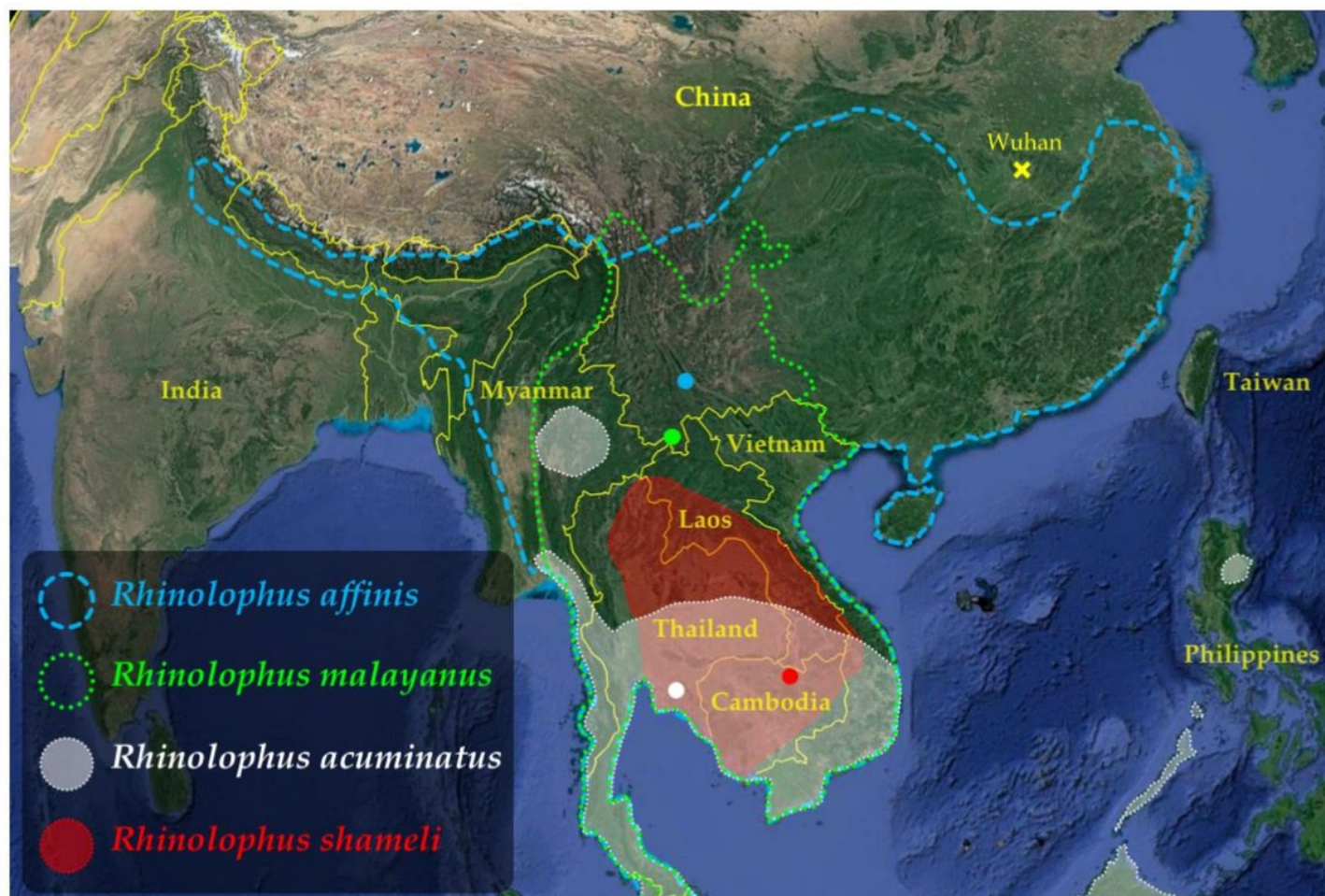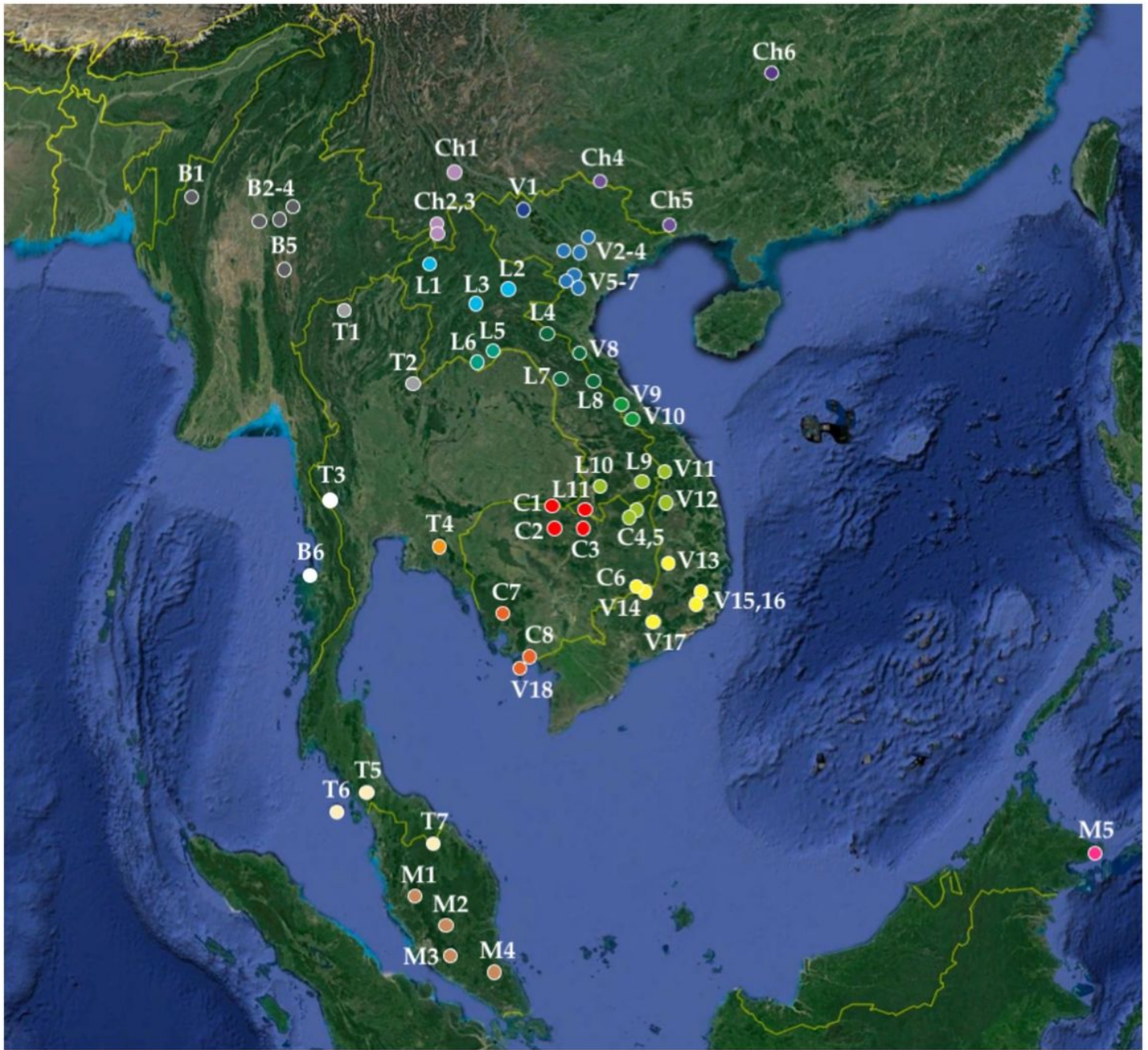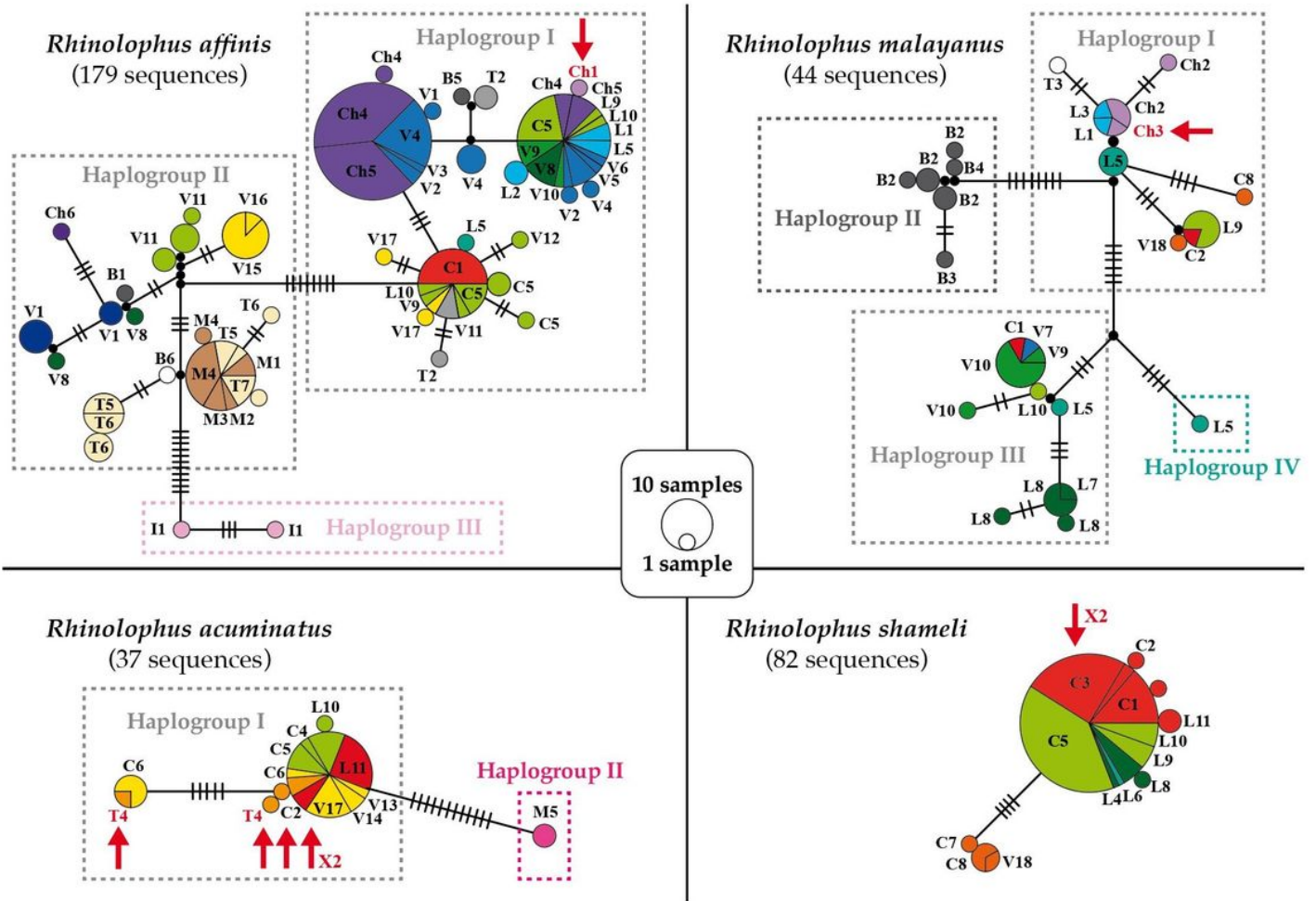
compliance with the ARRIVE guidelines. 517

Figures

Figure 1

Geographic distribution of the four Rhinolophus species found positive for viruses closely related to SARS-CoV-2 in southern China and Southeast Asia. Map from Google Earth US Dept of State Geographer © 2020 Google – Image Landsat / Copernicus - Data SIO, NOAA, U.S. Navy, NGA, GEBCO. For each of the four Rhinolophus species, the geographic distribution was extracted from the IUCN website [11]. The coloured dots show the four geographic locations where bats found positive for SCoV2rCs were collected Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.
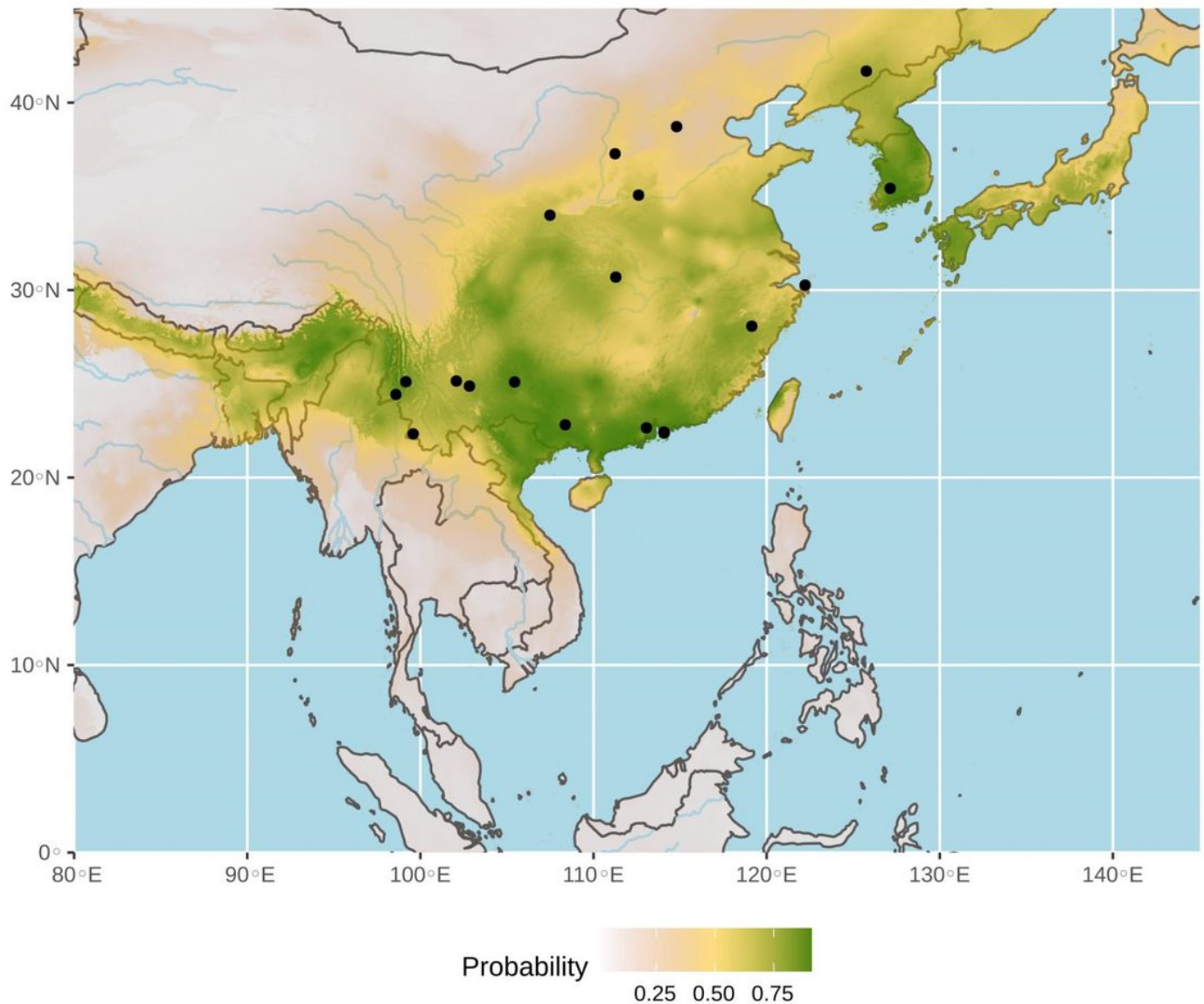
Figure 2

Geographic localities analysed in this study. The codes used for the countries are the following: B (Myanmar), C (Cambodia), Ch (China), I (Indonesia), L (Laos), M (Malaysia), T (Thailand), and V (Vietnam). Different regions were highlighted using colours. Map from Google Earth US Dept of State Geographer © 2020 Google – Image Landsat / Copernicus - Data SIO, NOAA, U.S. Navy, NGA, GEBCO. The locality names and GPS coordinates are provided in online supplementary table S1 Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.
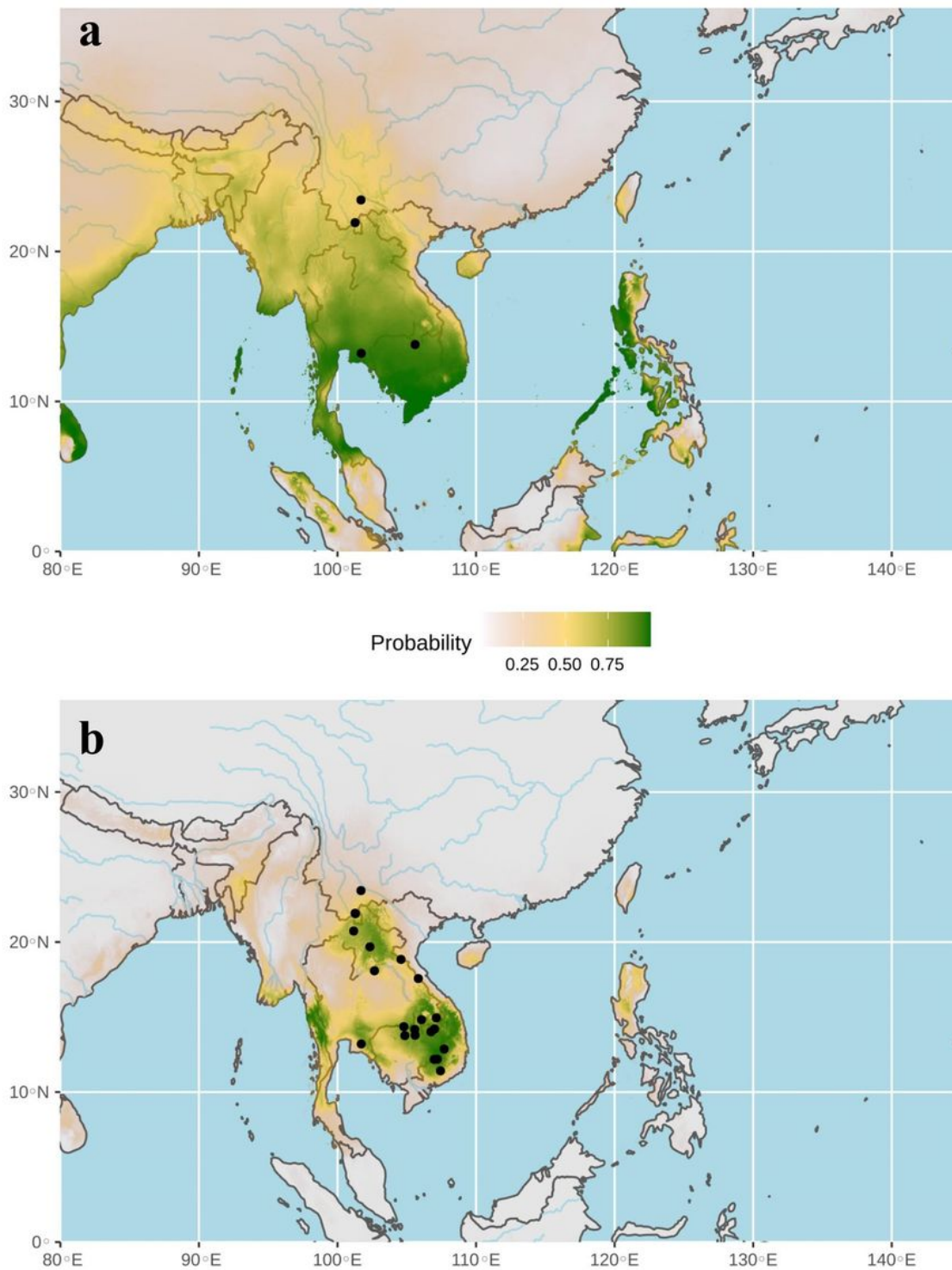
Figure 3

Haplotype networks based on CO1 sequences of the four Rhinolophus species found positive for viruses closely related to SARS-CoV-2 (SCoV2rCs). The networks were constructed with the median joining method available in PopART 1.5 [13]. The codes used for the countries are the following: B (Myanmar), C (Cambodia), Ch (China), I (Indonesia), L (Laos), M (Malaysia), T (Thailand), and V (Vietnam). Colours indicate the geographic origin of haplotypes according to Fig. 2 (see online supplementary table S1). The circles indicate haplotypes separated by at least one mutation. The black lines on the branches show the number of mutations ≥ 2. Black circles represent missing haplotypes. Circle size is proportional to the number of haplotypes. Haplogroups separated by more than seven mutations (pairwise nucleotide distances > 1%) are highlighted by dotted lines. The red arrows show the positions of the nine bats found positive for SCoV2rCs.

Figure 4

Ecological niche of bat viruses related to SARS-CoV (SCoVrCs). The geographic distribution of suitable environments was predicted using the Maxent algorithm in ENMTools (see Methods for details). AUC = 0.81. Black circles indicate localities used to build the distribution model (see geographic coordinates in online supplementary table S2). Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 5

Ecological niches of bat viruses closely related to SARS-CoV-2 (SCoV2rCs) predicted using 4 points (a) (AUC = 0.58) and 21 points (b) (AUC = 0.96). The geographic distributions of suitable environments were predicted using the Maxent algorithm in ENMTools (see Methods for details). Black circles indicate localities used to build the distribution model (see geographic coordinates in online supplementary table S1). Note: The designations employed and the presentation of the material on this map do not imply the

expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.
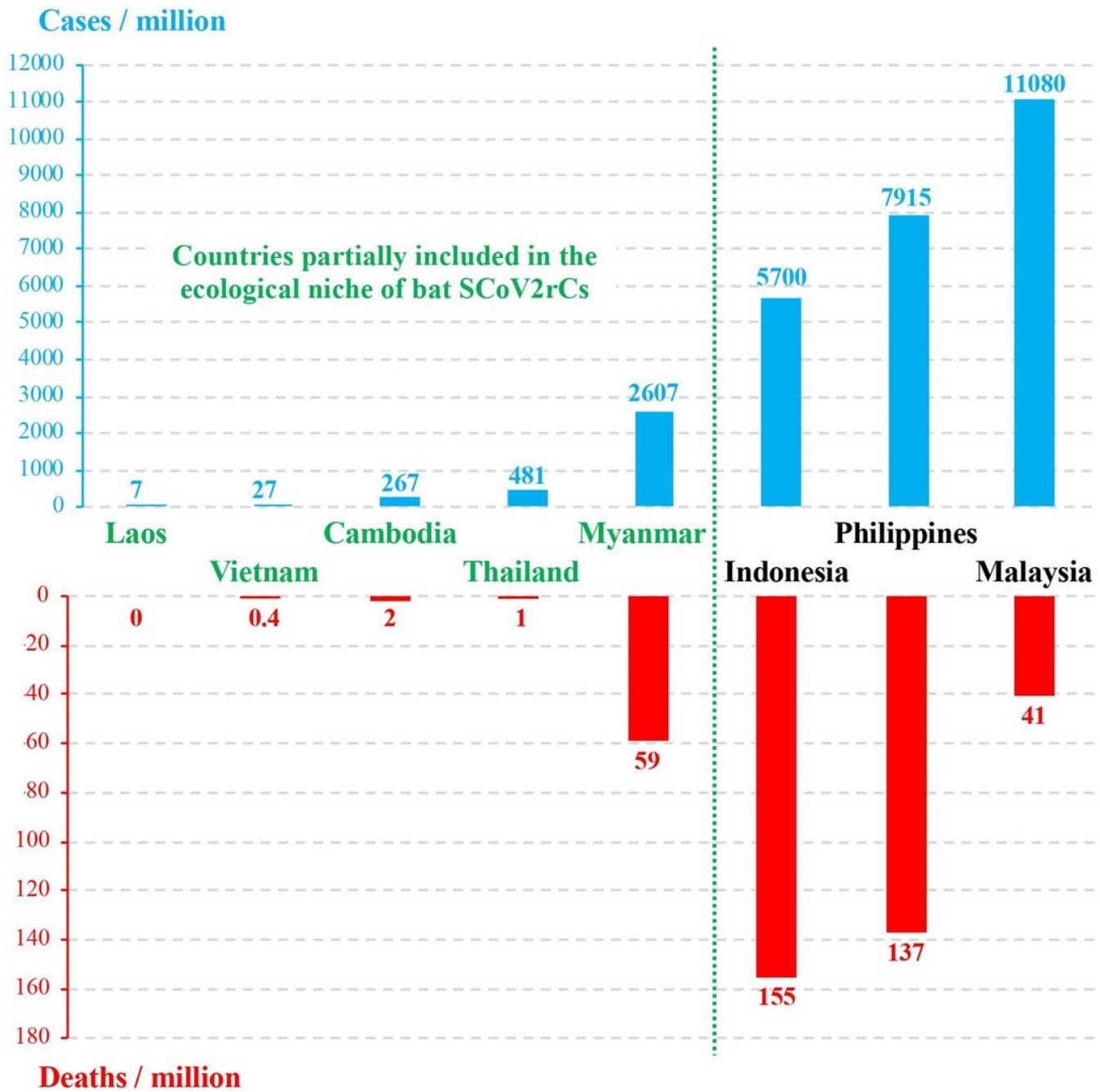

Figure 6

Number of Covid-19 patients per million inhabitants (in blue) and deaths per million inhabitants (in red) for the different countries of Southeast Asia. Data extracted from the Worldometers website [2] on April 12, 2021

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.pdf