

1 **Title:** SpeciesGeoCoder: Fast categorisation of species occurrences for analyses of biodiversity,
2 biogeography, ecology and evolution

3
4 Authors: Mats Töpel^{1,2,*}, Maria Fernanda Calió^{1,3,4}, Alexander Zizka¹, Ruud Scharn¹, Daniele
5 Silvestro¹, Alexandre Antonelli^{1,5,*}

6
7 ¹University of Gothenburg, Department of Biological and Environmental Sciences, Box 461, SE-
8 405 30 Göteborg, Sweden

9 ²Bioinformatics Infrastructure for Life Sciences, www.bils.se

10 ³Universidade de São Paulo, Instituto de Biociências, Departamento de Botânica, Rua do Matão,
11 277 - Cidade Universitária, CEP: 05508-090, São Paulo, SP, Brazil

12 ⁴Universidade Federal de São Carlos, Departamento de Botânica, CEP: 13565-905, São Carlos,
13 SP, Brazil

14 ⁵Gothenburg Botanical Garden, Carl Skottsbergs gata 22A, SE-41319, Göteborg, Sweden

15

16 *Corresponding authors; mats.topel@bioenv.gu.se, alexandre.antonelli@bioenv.gu.se

17

18 **Abstract:** Understanding the patterns and processes underlying the uneven distribution of
19 biodiversity across space and time constitutes a major scientific challenge in evolutionary
20 biology. With rapidly accumulating species occurrence data, there is an increasing need for
21 making the process of coding species into operational units for biogeographic and evolutionary
22 analyses faster, automated, transparent and reproducible. Here we present SpeciesGeoCoder, a
23 free software package written in Python and R, that allows for easy coding of species into user-
24 defined areas. These areas may be of any size and be purely geographical (i.e., polygons) such as
25 political units, conservation areas, biomes, islands, biodiversity hotspots, and areas of endemism,
26 but may also include altitudinal ranges. This flexibility allows scoring species into complex
27 categories, such as those encountered in topographically and ecologically heterogeneous
28 landscapes. In addition, SpeciesGeoCoder can be used to facilitate sorting and cleaning of
29 occurrence data. The various outputs of SpeciesGeoCoder include quantitative biodiversity
30 statistics, global and local distribution maps, and NEXUS files that can be directly used in many
31 phylogeny-based applications for ancestral state reconstruction, investigations on biome

32 evolution, and diversification rate analyses. Our simulations indicate that even datasets
33 containing hundreds of millions of records can be analysed in relatively short time using a
34 regular desktop computer. We exemplify the use of our program through two contrasting
35 examples: *i*) inferring historical dispersal of birds across the Isthmus of Panama, separating
36 lowland *vs.* montane species and optimising the results onto a species-level, dated phylogeny;
37 and *ii*) exploring seasonal variations in the occurrence of 10 GPS-tracked individuals of moose
38 (*Alces alces*) over one year in northern Sweden. These analyses show that SpeciesGeoCoder
39 allows an easy, flexible and fast categorisation of species distribution data for various analyses in
40 ecology and evolution, with potential use at different spatial, taxonomic and temporal scales.

41

42 **Key words:** Biodiversity – Biogeography – Bioinformatics – Species distributions – Evolution.

43 **Introduction**

44 Species distributions provide the basic knowledge for biodiversity research, allowing us to
45 understand their environmental requirements, biogeographic history, and expected resilience to
46 climate change. However, analysing the distribution of the world's estimated 8.7 million species
47 (Mora, C., Tittensor, D.P., et al. 2011) remains a major scientific challenge.

48
49 There are now over half a million species occurrences worldwide freely available through the
50 Global Biodiversity Information Facility (GBIF; <http://www.gbif.org>) and their partners, of
51 which the majority (about 446 million) are geo-referenced, i.e. provided with a latitude and
52 longitude. These numbers are steadily increasing thanks to new agreements on data sharing, on-
53 going digitalisation programmes, the work of active field biologists and amateurs combined with
54 an increasing awareness of the need to provide rich metadata with biological collections and
55 sequences (Hyde KD, U.D., Manamgoda DS, Tedersoo L, Nilsson RH. 2013), and tools that
56 enable automated geo-referencing of older museum specimens (Guralnick, R.P., Wieczorek, J.,
57 et al. 2006, Garcia-Milagros, E. and Funk, V.A. 2010). Publicly available species occurrences
58 represent an enormous data source for biodiversity research, but are as yet poorly exploited due
59 to two main factors: *i*) general scepticism concerning the quality of records available, in terms of
60 species identification and precise coordinates (Hjarding, A., Tolley, K.A., et al. 2014), and *ii*)
61 demonstrated taxonomic, geographic, and temporal biases (Boakes, E.H., McGowan, P.J., et al.
62 2010). Whereas these biases are slowly being compensated by data growth, improving quality
63 typically relies on expert curation and targeted evaluations (e.g. <http://www.iucnredlist.org/>) as
64 well as tools for massive data cleaning (e.g. through workflows at the Biodiversity Virtual e-
65 Laboratory, <http://www.biovel.eu>).

66
67 Research using species distribution data has been further hampered by a lack of tools that allow
68 for classification of raw occurrences into discrete categories. This is often a crucial step in many
69 evolutionary and biogeographic analyses, since such categories can then be used in connection
70 with e.g. a phylogeny for ancestral range reconstructions (Pagel, M., Meade, A., et al. 2004, Ree,
71 R.H. and Smith, S.A. 2008, Matzke, N.J. 2013) and for the estimation of area-dependent
72 inferences of diversification rates (Goldberg, E.E., Lancaster, L.T., et al. 2011, Silvestro, D.,
73 Schnitzler, J., et al. 2011, FitzJohn, R.G. 2012). Species categorisation into commonly

74 recognised areas such as eco-regions and realms (Olson, D.M., Dinerstein, E., et al. 2001, Abell,
75 R., Thieme, M.L., et al. 2008, Holt, B.G., Lessard, J.-P., et al. 2013) may also reveal patterns of
76 biodiversity and distribution at a large scale, facilitating the identification of regions with
77 outstanding levels of species richness and endemism – central to the concept of Biodiversity
78 Hotspots (Myers, N., Mittermeller, R.A., et al. 2000).

79
80 Rapidly increasing amounts of species occurrence data, the need to classify species into discrete
81 areas in an automated, reproducible and transparent way tailored to evolutionary biologists, have
82 led us to develop **SpeciesGeoCoder**.

83

84 **Description**

85 SpeciesGeoCoder is written in Python (<https://www.python.org/>) and makes use of the R
86 software environment (<http://www.r-project.org/>) for statistics and plotting of results. It runs on
87 all platforms for which these programs are available, including Windows, MacOS X,
88 GNU/Linux and BSD, and can be run either through the command-line or through a graphical
89 user interface (GUI). The GUI provides researchers that are not familiar with using command-
90 line programs, an easy-to-use set of method for analysing large amounts of geo-referenced data.
91 The source code consists of a set of python modules, and the analyses of GeoTIFF files are done
92 using the GDAL python bindings (<http://www.gdal.org/>) for fast execution. The basic workflow
93 describing the package is illustrated in Figure 1.

94

95 SpeciesGeoCoder works as follows:

- 96 1. The user provides input data in two files: A) species occurrence data (including species
97 names, latitude and longitude, in tab-delimited text format [see example files distributed with
98 the program] or in the format directly provided by GBIF); and B) a text file defining the
99 areas to be used for coding, i.e. a list of polygon names and coordinates for each of its edges,
100 and (optionally) an altitudinal range (e.g., between 500 – 1000 meters above sea level). The
101 polygons can be easily designed through various GIS tools (e.g. the freely available program
102 QGIS, <http://qgis.osgeo.org>). If an altitudinal range is provided, additional altitudinal data
103 files for the regions covered and at the resolution desired (in GeoTIFF format) are needed.

- 104 This type of data can be freely downloaded from various on-line resources (see
105 [SpeciesGeoCoder's wiki](#) for a list of repositories).
- 106 2. SpeciesGeoCoder loops through all samples in the input file, counting the presences for each
107 species in each polygon.
 - 108 3. The default output is a NEXUS file (Maddison, D.R., Swofford, D.L., et al. 1997) containing
109 a data matrix with all analysed species and their presence or absence in each area coded as '1'
110 or '0', respectively. 'Presence' requires by default at least a single occurrence in an area, but
111 may be set to require a number above a user-defined threshold (either an absolute number, a
112 percentage of the occurrences, or both). Alternatively, the user may ask for a more complex
113 output including the number of occurrences in each polygon, which could aid the
114 identification of outliers that could require further verification; this information appears as
115 comments in the NEXUS files. This file can be directly analysed in programmes that handle
116 the NEXUS input format, such as Mesquite (Maddison, W.P. and Maddison, D.R. 2009) and
117 most phylogenetic packages written in R, such as APE (Paradis, E., Claude, J., et al. 2004),
118 GEIGER (Harmon, L.J., Weir, J.T., et al. 2008), and Diversitree (FitzJohn, R.G. 2012), as
119 well as others written in Python, such as BayesRates (Silvestro, D., Schnitzler, J., et al. 2011).
 - 120 4. The second (optional) result is a series of summary statistics and distribution maps. These
121 include multiple PDF documents with bar charts as graphical representations of the number
122 of species per area, the number of occurrences per species per area, and the relative
123 occurrence per area for each species. The summary tables used for the graphical output are
124 also provided as tab-delimited text files to facilitate downstream analyses. The distribution
125 maps plot all occurrence points and the areas included in the analyses, coded at the species
126 level. In addition, for datasets including less than 40 species, a co-existence matrix for each
127 area is calculated and visualized as a heat plot.
 - 128 5. The third (optional) result is a series of plots summarising the historical movements (range
129 expansions or dispersals) of lineages between all pairs of user-defined areas, based on a dated
130 phylogeny (or a sample of trees, to account for topological and divergence time uncertainties).
131 These plots are computed with novel scripts that employ available packages in R, stochastic
132 mapping of shifts in transitions along branches, and the computation of absolute as well as
133 relative numbers of dispersals through time, i.e. corrected by the number of lineages
134 (Silvestro, D. 2012, Fernández-Mendoza, F. and Printzen, C. 2013). Other methods for the

135 reconstruction of ancestral areas, e.g. incorporating the possibility of vicariance and founding
136 effects (Ree, R.H. and Smith, S.A. 2008, Matzke, N.J. 2013), may be readily used based on
137 the output of SpeciesGeoCoder but are not implemented in the first release of this package.

138

139 **Benchmark**

140 We evaluated the performance and scalability of SpeciesGeoCoder through a series of
141 simulations. We focused on how computing time is determined by three key variables: *i*) the
142 number of geo-referenced occurrences, *ii*) the number of polygons, and *iii*) the complexity of the
143 polygons, measured by their number of edges (corners). The occurrence dataset (i) was simulated
144 as a set of points evenly distributed longitudinally and latitudinally (see example in Fig. S1). The
145 polygon dataset (ii) was generated in a similar way, by creating a grid of square polygons sharing
146 two corners with each of its neighbouring polygons (Fig. S2). The edge dataset (iii) was
147 generated by creating one square polygon and successively adding corners equally distributed
148 over its perimeter (Fig. S3). The simulations were performed with a logarithmically increase in
149 the number of occurrences, polygons, and polygon edges, ranging between 10^1 and 10^8 .

150

151 Figure 2 summarises the results of these simulations, and shows that there is a linear increase in
152 computing time in relation to the three variables examined. These results also show that
153 SpeciesGeoCoder can handle vast amounts of data within feasible time using regular computer
154 hardware.

155

156 **Biological Examples**

157 *Bird biogeography*. We inferred the historical dispersal of montane and lowland bird lineages
158 through time across the Central American Seaway, which separated North and South America
159 for millions of years until the emergence of the Isthmus of Panama. First, we downloaded the full
160 occurrence dataset including c. 10,000 species and 200,000,000 records obtained from
161 <http://www.ebird.org> (eBird 2013). We then used SpeciesGeoCoder to identify all records found
162 outside the American continent, as well as all those found north of Mexico, which we then
163 excluded from further analyses. Species pertaining to the remaining records were coded as
164 occurring in Central America, South America, or both. Rather than using political boundaries,
165 we defined the border between these two areas following the Uramita fault (Montes, C., Bayona,

166 G., et al. 2012) that separates the South American and Panamanian geological plates (polygons
167 shown in Fig. S4). We created two operational units from each polygon, one with the additional
168 condition of only including records occurring below 1000 meters above sea level, and the other
169 for records occurring above this altitudinal threshold, following the same categorisation as Weir
170 (2006). We then reconstructed ancestral areas onto the species-level dated phylogeny of birds
171 from Jetz et al. (2012), using stochastic mapping to reconstruct the historical dispersal of
172 lineages through time among these four operational units. We calculated both the total (absolute)
173 as well as the relative (in proportion to the number of lineages) number of dispersals between
174 each pair of areas, using bins of 10 million years. Due to issues with synonymy, species concepts,
175 and/or lack of geo-referenced data (issues which we could not readily solve computationally), the
176 analyses dataset was reduced to include 4350 species.

177
178 Our results (Fig. 3) suggest that dispersals between the lowlands of South and Central America
179 occurred considerably more frequently (c. 2–4 times) than dispersals between the highlands of
180 those landmasses. There were no major differences in directionality of dispersals, except for the
181 last time bin considered (0–10 Ma), when northwards dispersals dominated. This supports the
182 conclusion by Weir et al. (2009) that birds mainly followed an opposite route during the Great
183 American Biotic Interchange as compared to mammals, which migrated mostly southwards
184 (Stehli, F.G. and Webb, S.D. 1985). The rate of dispersals increased for all categories in the most
185 recent time bin, reflecting the full emergence of the Panama Isthmus.

186
187 *Seasonal behaviour in Swedish moose.* We also tested the functionality of SpeciesGeoCoder at a
188 much lower taxonomic, spatial and temporal level, by analysing a dataset of ten GPS-tracked
189 individuals of moose (*Alces alces*) over one year in northern Sweden (c. 80,000 records). The
190 position data were retrieved from the Wireless Remote Animal Monitoring database system for
191 data validation and management (Dettki, H., Ericsson, G., et al. 2013). We created five arbitrary
192 polygons, as if they would correspond to e.g. planned national parks or areas for urban
193 development, where knowledge on wildlife activity could play a role in decision making. We
194 calculated the co-occurrence of individuals in each of those polygons during four seasons,
195 correcting for season length (which followed a local meteorological classification, resulting in
196 seasons of different numbers of days: spring 60, summer 48, autumn 71, winter 186).

197

198 The results of the moose analysis are plotted together with the original data points in Fig. 4.
199 Moose individuals spent most of the time in the north-western polygon, despite their being
200 concentrated to a narrow strip. Co-occurrence patterns varied among seasons, with individuals
201 being more clustered during the spring (when calves are usually born) than in the other seasons,
202 especially in the winter (when food is more scarce).

203

204 **Further prospects**

205 Beyond the examples provided here, the output obtained by SpeciesGeoCoder could be readily
206 used for e.g. calculating measures of alpha, beta and gamma diversity; the identification of
207 neglected areas for conservation; and providing real-time detection of GPS-tagged animals
208 entering and leaving protected areas. In addition, the visualisation and coding of species into
209 areas may greatly facilitate cleaning up occurrence databases, by enabling the identification of
210 outliers that may require additional examination and possible exclusion from downstream
211 analyses.

212

213 **Availability**

214 SpeciesGeoCoder is free software available under the GPL3 licence from
215 <https://github.com/mtop/speciesgeocoder/releases>. The release and its associated Wiki page
216 (<https://github.com/mtop/speciesgeocoder/wiki>) include installation instructions for Mac OSX,
217 Gnu/Linux (and other UNIX-like systems) and Windows, bundled installation packages,
218 example files, tutorials, a list of online repositories for GeoTIFF data, pre-defined polygons, and
219 other useful links. A graphical user interface that provides the core functions of the program is
220 available at <http://sourceforge.net/projects/speciesgeocodergui>. All pages can be accessed from
221 <http://antonelli-lab.net>.

222

223 **Author contributions**

224 A.A. and M.T. initiated the project; M.F.C. and R.S. produced the tutorial; A.Z. developed the R
225 features; D.S. implemented the scripts for stochastic mapping; M.F.C., A.Z., R.S. and D.S.
226 performed the simulations and analysed the biological data; M.T. led the development of the
227 code; A.A. led the writing with contributions from all authors. All authors participated in the

228 design and implementation of the software, the production of figures and associated material,
229 and approved the final version of this manuscript.

230

231 **Acknowledgments**

232 We thank Rosemeri Morokawa, Alexander Walther, students and colleagues at our labs for
233 fruitful discussions and software testing, and Fernanda Carvalho, Henrik Nilsson and Ricardo
234 Sawaya for comments on this manuscript. We also acknowledge Holger Dettki and Göran
235 Ericsson for providing the moose data. Finally, we thank Paula Töpel for designing the
236 SpeciesGeoCoder logo. Funding was provided by the Swedish Research Council (B0569601)
237 and the European Research Council under the European Union's Seventh Framework
238 Programme (FP/2007-2013, ERC Grant Agreement n. 331024) to A.A., from Carl Tryggers
239 stiftelse ([CTS 11:479](#), [CTS 12:507](#)) to M.T., and from Fundação de Amparo à Pesquisa do
240 Estado de São Paulo (FAPESP 2009/52161-2, 2013/10262-2) to M.F.C. The code was developed
241 with support from the Bioinformatics Infrastructure for Life Sciences (<http://www.bils.se>) and
242 tested and benchmarked on the bioinformatics computer cluster Albiorix at the Department of
243 Biological and Environmental Sciences, University of Gothenburg.

244

245 **References**

246 Abell R, Thieme ML, Revenga C, Bryer M, Kottelat M, Bogutskaya N, Coad B, Mandrak N,
247 Balderas SC, Bussing W. 2008. Freshwater ecoregions of the world: a new map of biogeographic
248 units for freshwater biodiversity conservation. *Bioscience*, 58:403-414.

249 Boakes EH, McGowan PJ, Fuller RA, Chang-qing D, Clark NE, O'Connor K, Mace GM. 2010.
250 Distorted views of biodiversity: spatial and temporal bias in species occurrence data. *PLoS Biol.*,
251 8:e1000385.

252 Dettki H, Ericsson G, Giles T, Norrsken-Ericsson M. 2013. Wireless Remote Animal Monitoring
253 (WRAM) - A new international database e-infrastructure for telemetry sensor data from fish and
254 wildlife. *Proceedings Etc 2012: Convention for Telemetry, Test Instrumentation and Telecontrol*
255 (Eds. The European Society of Telemetry). Books on Demand, pp. 292, ISBN: 978-3-7322-
256 5646-4.:247-256.

257 eBird. 2013. Version: EBD_relAug-2013. Cornell Lab of Ornithology, Ithaca, New York.
258 August 2013.

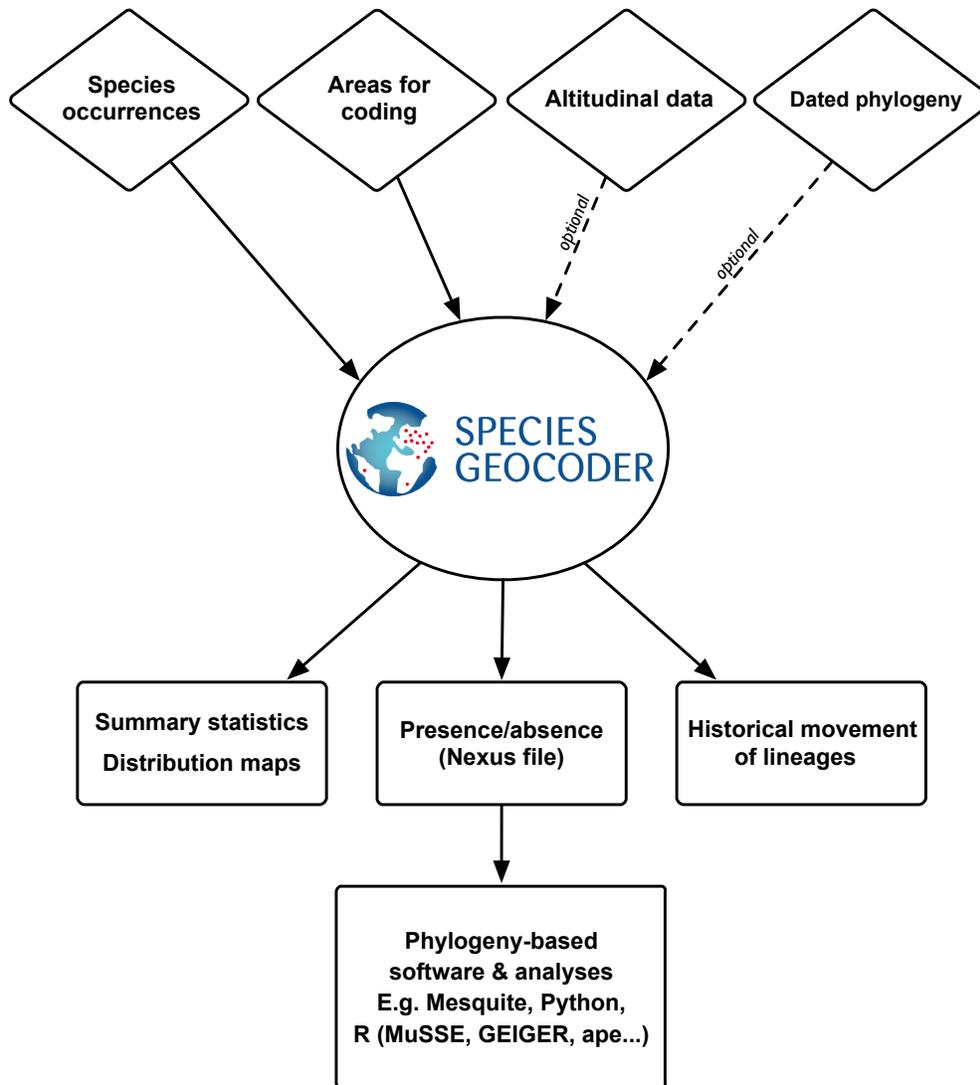
259 Fernández-Mendoza F, Printzen C. 2013. Pleistocene expansion of the bipolar lichen *Cetraria*
260 *aculeata* into the Southern hemisphere. *Mol. Ecol.*, 22:1961-1983.

- 261 FitzJohn RG. 2012. Diversitree: comparative phylogenetic analyses of diversification in R.
262 *Methods in Ecology and Evolution*, 3:1084-1092.
- 263 Garcia-Milagros E, Funk VA. 2010. Improving the use of information from museum specimens:
264 Using Google Earth© to georeference Guiana Shield specimens in the US National Herbarium.
265 *Frontiers of Biogeography*, 2.
- 266 Goldberg EE, Lancaster LT, Ree RH. 2011. Phylogenetic inference of reciprocal effects between
267 geographic range evolution and diversification. *Syst. Biol.*, 60:451-465.
- 268 Guralnick RP, Wieczorek J, Beaman R, Hijmans RJ, the BioGeomancer Working G. 2006.
269 BioGeomancer: Automated Georeferencing to Map the World's Biodiversity Data. *PLoS Biol*,
270 4:e381.
- 271 Harmon LJ, Weir JT, Brock CD, Glor RE, Challenger W. 2008. GEIGER: investigating
272 evolutionary radiations. *Bioinformatics*, 24:129-131.
- 273 Hjarding A, Tolley KA, Burgess ND. 2014. Red List assessments of East African chameleons: a
274 case study of why we need experts. *Oryx*, FirstView:1-7.
- 275 Holt BG, Lessard J-P, Borregaard MK, Fritz SA, Araújo MB, Dimitrov D, Fabre P-H, Graham
276 CH, Graves GR, Jønsson KA. 2013. An update of Wallace's zoogeographic regions of the world.
277 *Science*, 339:74-78.
- 278 Hyde KD UD, Manamgoda DS, Tedersoo L, Nilsson RH. 2013. Incorporating molecular data in
279 fungal systematics: a guide for aspiring researchers. *Current Research in Environmental and*
280 *Applied Mycology*, 3:1-32.
- 281 Jetz W, Thomas G, Joy J, Hartmann K, Mooers A. 2012. The global diversity of birds in space
282 and time. *Nature*, 491:444-448.
- 283 Maddison DR, Swofford DL, Maddison WP. 1997. NEXUS: an extensible file format for
284 systematic information. *Syst. Biol.*, 46:590-621.
- 285 Maddison WP, Maddison DR. 2009. Mesquite: a modular system for evolutionary analysis.
286 Version 2.71 <http://mesquiteproject.org>.
- 287 Matzke NJ. 2013. BioGeoBEARS: BioGeography with Bayesian (and Likelihood) Evolutionary
288 Analysis in R Scripts. University of California, Berkeley, Berkeley, CA.
- 289 Montes C, Bayona G, Cardona A, Buchs DM, Silva CA, Morón S, Hoyos N, Ramírez DA,
290 Jaramillo CA, Valencia V. 2012. Arc-continent collision and orocline formation: Closing of the
291 Central American seaway. *J. Geophys. Res.*, 117:B04105.
- 292 Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B. 2011. How Many Species Are There on
293 Earth and in the Ocean? *PLoS Biol*, 9:e1001127.

- 294 Myers N, Mittermeyer RA, Mittermeyer CG, Da Fonseca GAB, Kent J. 2000. Biodiversity
295 hotspots for conservation priorities. *Nature*, 403:853-858.
- 296 Olson DM, Dinerstein E, Wikramanayake ED, Burgess ND, Powell GV, Underwood EC,
297 D'amico JA, Itoua I, Strand HE, Morrison JC. 2001. Terrestrial Ecoregions of the World: A New
298 Map of Life on Earth. *Bioscience*, 51:933-938.
- 299 Pagel M, Meade A, Barker D. 2004. Bayesian estimation of ancestral character states on
300 phylogenies. *Syst. Biol.*, 53:673-684.
- 301 Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution in R
302 language. *Bioinformatics*, 20:289-290.
- 303 Ree RH, Smith SA. 2008. Maximum likelihood inference of geographic range evolution by
304 dispersal, local extinction, and cladogenesis. *Syst. Biol.*, 57:4-14.
- 305 Silvestro D. 2012. Diversification in time and space. Methodological advancement and case
306 studies from the Neotropical plant family Bromeliaceae. Johann-Wolfgang-Goethe-University,
307 Frankfurt am Main, Germany.
- 308 Silvestro D, Schnitzler J, Zizka G. 2011. A Bayesian framework to estimate diversification rates
309 and their variation through time and space. *BMC Evol. Biol.*, 11:311.
- 310 Stehli FG, Webb SD. 1985. The Great American Biotic Interchange.
- 311 Weir JT. 2006. Divergent timing and patterns of species accumulation in lowland and highland
312 neotropical birds. *Evolution*, 60:842-855.
- 313 Weir JT, Bermingham E, Schluter D. 2009. The Great American Biotic Interchange in birds.
314 *Proc. Natl. Acad. Sci. USA*, 106:21737-21742.
- 315
- 316

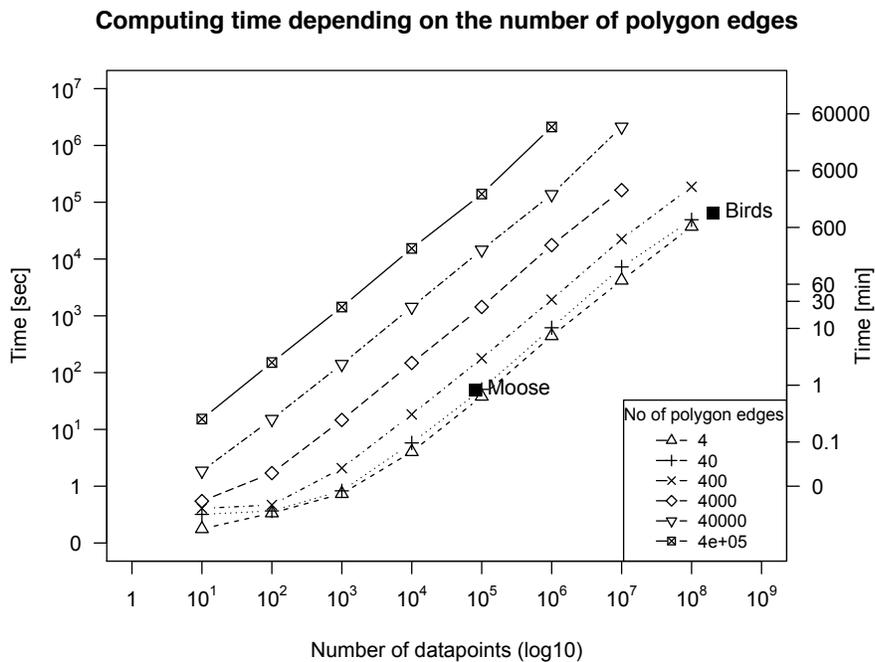
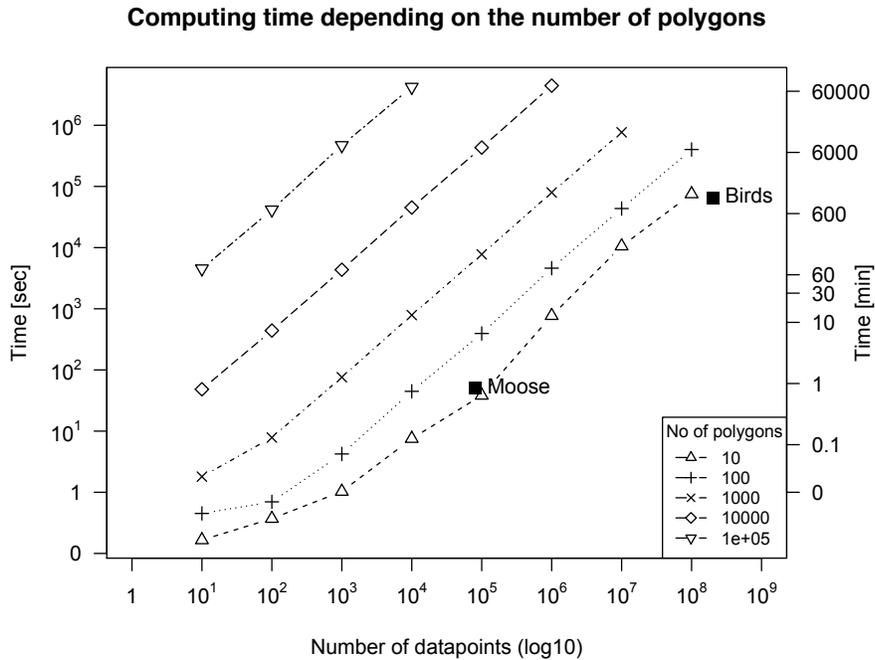
317 **Figure legends**

318



319

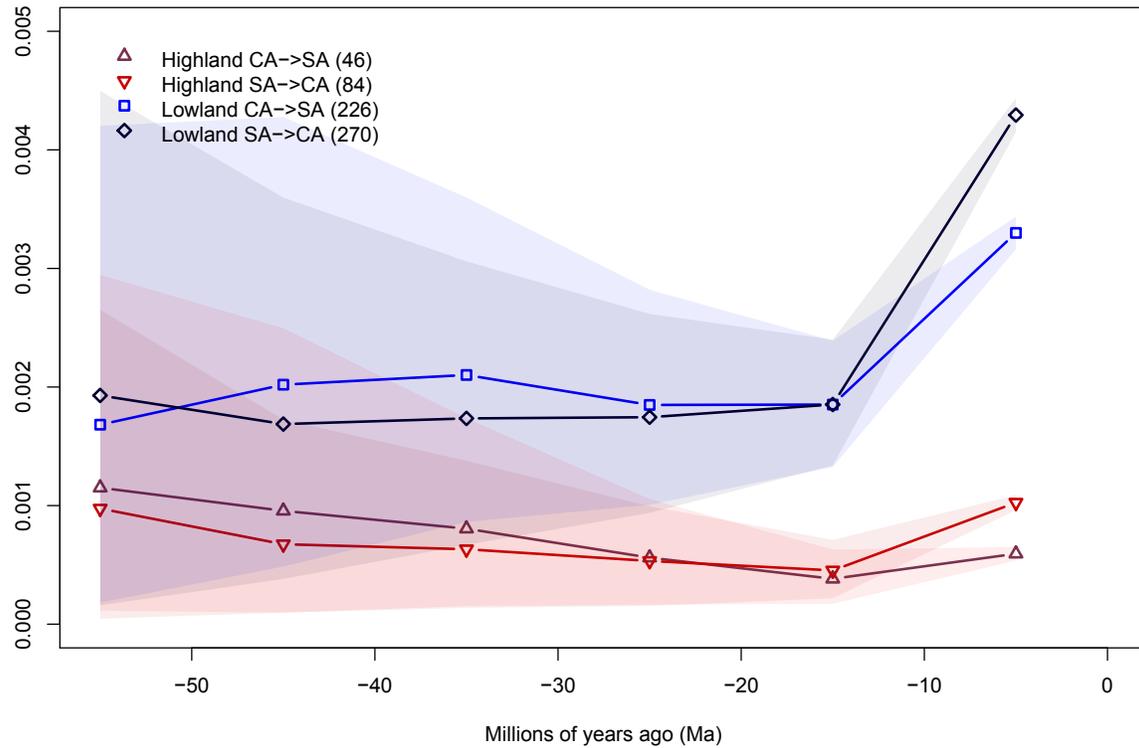
320 **Figure 1.** Simplified workflow of the SpeciesGeoCoder package.



321

322 **Figure 2.** Computational time in relation to the increase in the number of polygons, polygon
 323 complexity (number of edges) and number of species occurrences, as estimated through
 324 simulations. As a comparison to empirical data, the dot denoted ‘Birds’ corresponds the coding
 325 of all c. 200,000,000 bird occurrences available from ebird.org; while ‘Moose’ corresponds to
 326 the coding of c. 80000 records, as described in the text.

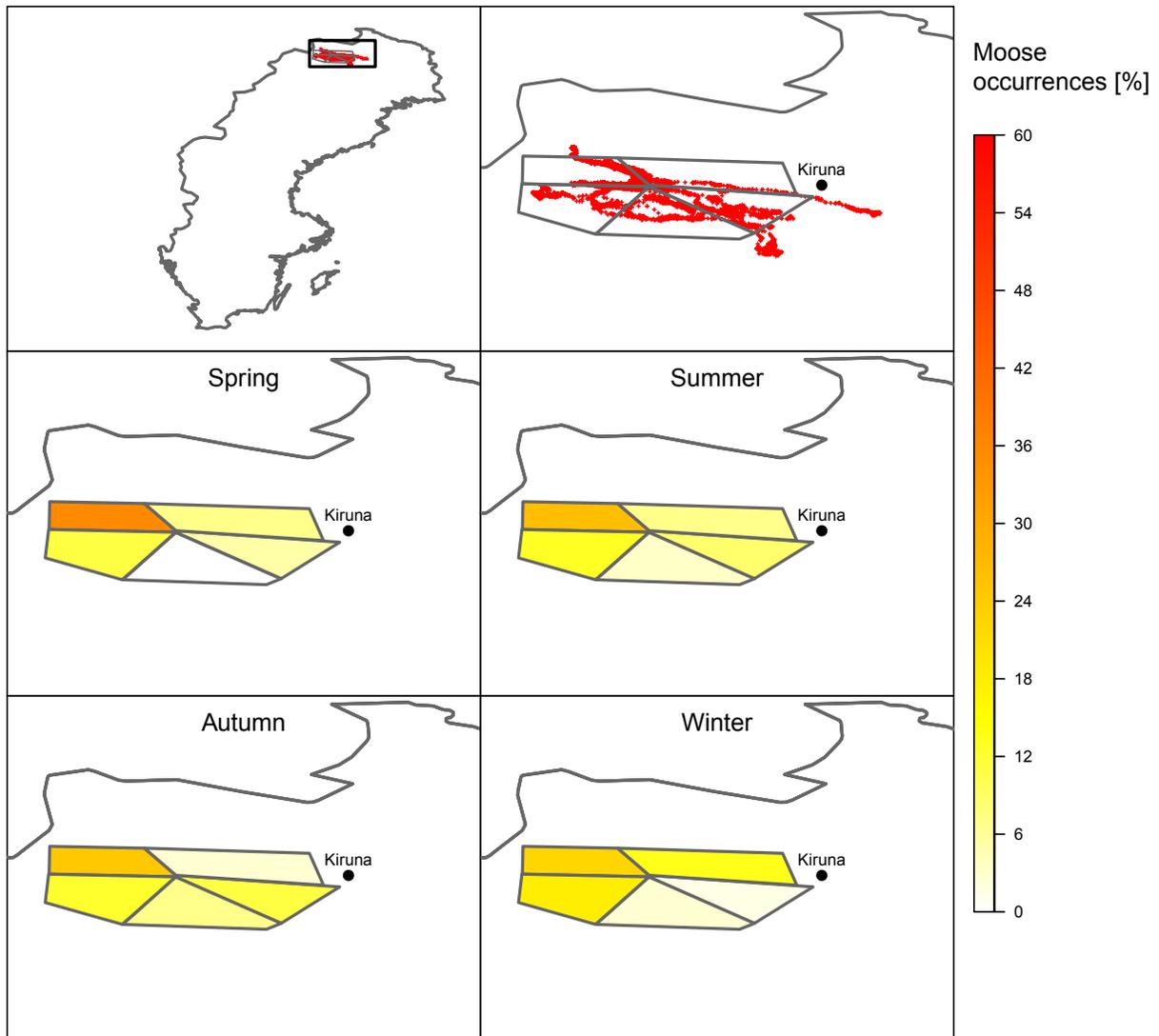
Historical events of bird dispersals between Central and South America



327

328 **Figure 3.** Relative frequency (Y-axis) of historical range movements (dispersals) of lowland and
329 highland bird lineages between South America (SA) and Central America (CA), calculated per
330 10-million-year time bins. The total number of dispersal events inferred for our dataset is
331 indicated between brackets in the legend.

332



333

334 **Figure 4.** Spatial and seasonal variation of occurrences for 10 individuals of moose (*Alces alces*)
335 in northern Sweden, scaled to the number of days in each season. The numbers on the heat scale
336 correspond to hourly records obtained from the Wireless Remote Animal Monitoring database
337 system and span over one year.