

1 **Complementando inventarios biológicos con datos abiertos a través de un método**
2 **semiautomatizado**

3
4 **ABSTRACT**

5 **Introduction:** Surveys are the best way to document the biological richness of any given place,
6 but efforts to do so can be limited. We asked whether the species richness documented through
7 field work can be complemented with open data from digital platforms. **Objective:** To analyze
8 whether species occurrence data available in the *Global Biodiversity Information Facility* (GBIF)
9 can complement species lists reported in the literature that were obtained through field work.

10 **Methods:** We compared bird species lists published in peer-reviewed articles and those
11 generated with information from GBIF for the same study areas using the Sorensen dissimilarity
12 index (β_{sor}) and its two components (turnover β_{sim} and nestedness β_{sne}). **Results:** GBIF reported
13 more species ($n = 26 - 232$) than 14 of the 24 lists consulted, but fewer ($n = 12 - 114$) than the
14 other 10. The sum of species documented by both sources represents an increase between 1.6%
15 and 464% compared with the number of species in the articles alone, showing the potential for
16 complementation between both sources. β_{sor} ranged between 0.15 – 1, showing that there always
17 were differences between species lists. In no instance $\beta_{\text{sim}} = 0$ or $\beta_{\text{sne}} = 1$, indicating that one
18 source of information always listed species absent in the other source and that species nestedness
19 between sources of information was never total. **Conclusion:** Species occurrences available in
20 GBIF can complement the information collected in the field but this in turn can also complement
21 the records of GBIF, so we highlight the importance of continuing documenting the biological
22 richness through field work, especially in little explored areas. We provide R code that can be
23 adapted to download and clean GBIF data for any biological group and place.

24
25 **Key words:** biodiversity; biodiversity informatics; biological inventories; citizen science; GBIF;
26 open data; species composition.

RESUMEN

27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53

Introducción: Los muestreos directos constituyen la mejor manera de documentar la riqueza biológica de un lugar, pero pueden verse limitados por diversas circunstancias. Nos preguntamos si la riqueza de especies documentada en campo puede ser complementada con datos de plataformas digitales de biodiversidad. **Objetivo:** Analizar si los datos de ocurrencia de especies disponibles en la *Global Biodiversity Information Facility* (GBIF) pueden complementar listados de especies creados con trabajo de campo y reportados en la literatura. **Métodos:** Comparamos listados de especies de aves disponibles en publicaciones académicas con aquellos generados usando información de GBIF para las mismas áreas de estudio utilizando el índice de disimilitud de Sorensen (β_{sor}) y sus dos componentes (recambio β_{sim} y anidamiento β_{sne}). **Resultados:** GBIF reporta más especies ($n = 26 - 232$) que 14 de 24 listados consultados, pero menos ($n = 12 - 114$) que los otros 10. La suma de especies documentadas entre fuentes representa un incremento de entre 1.6% y 464% comparada con el número de especies listadas solamente en los artículos, demostrando el potencial de complementación entre ambas fuentes. β_{sor} osciló entre 0.15 - 1, mostrando que siempre hubo diferencias entre los listados comparados. En ningún caso $\beta_{\text{sim}} = 0$, ni $\beta_{\text{sne}} = 1$, indicando que una fuente de información siempre listó especies ausentes en la otra y que el anidamiento de especies entre fuentes de información nunca fue total. **Conclusiones:** Los registros de ocurrencia de especies disponible en GBIF puede complementar la información recopilada en campo, pero ésta a su vez puede complementar los registros de GBIF, por lo que resaltamos la importancia de continuar documentando la riqueza biológica a través de trabajo de campo, especialmente en áreas poco exploradas. Proporcionamos código de R que puede ser adaptado para descargar y limpiar datos de GBIF de cualquier grupo biológico y lugar.

Palabras clave: Biodiversidad, ciencia ciudadana, composición de especies, datos abiertos, GBIF, informática de biodiversidad, inventarios biológicos.

Total de palabras: 8128

INTRODUCCIÓN

54
55 La obtención de conocimiento detallado de la riqueza biológica de una región requiere de
56 una inversión importante de tiempo, esfuerzo y recursos económicos (Escobar et al., 2009). Por
57 ejemplo, para estimar la riqueza avifaunística de lugares que son parte del sistema de migración
58 de aves Neártico - Neotropical, es necesario realizar observaciones durante al menos un año para
59 considerar la presencia temporal de especies migratorias. De manera similar, la baja detección de
60 especies elusivas en regiones megadiversas, aún sean residentes permanentes, hace necesaria la
61 realización de monitoreos de largo plazo. Por ejemplo, tres años de monitoreos mensuales de
62 murciélagos con redes de niebla y grabaciones acústicas en el Istmo de Tehuantepec fueron
63 insuficientes para caracterizar a toda la comunidad, ya que algunas especies sólo fueron
64 registradas como cadáveres al pie de aerogeneradores (Cabrera-Cruz et al., 2023).

65 Las curvas de acumulación de especies, aunadas a estimadores de riqueza, son
66 herramientas que permiten estimar tanto la completitud de inventarios biológicos como el
67 número adicional de especies que se podría esperar encontrar en un área de estudio (Colwell et
68 al., 2004; Magurran, 2004). Con frecuencia, estas herramientas indican que los inventarios están
69 incompletos a pesar de esfuerzos de muestreo importantes. Por ejemplo, monitoreos de aves
70 mensuales durante casi dos años en tres sitios al norte de México arrojaron valores de
71 completitud de 80% (Castillo-Muñoz y Guzmán-Hernández, 2021). Si bien se puede considerar
72 que un inventario con ese nivel de completitud es representativo de la diversidad avifaunística
73 del área de estudio, también ejemplifica el nivel de esfuerzo necesario para lograr dicha
74 representatividad.

75 No siempre es posible o necesario realizar un esfuerzo de muestreo intensivo. Diversas
76 circunstancias, como el financiamiento limitado, la dificultad en conseguir permisos o equipo, o
77 condiciones de inseguridad, entre otras, pueden limitar la cantidad de muestreos realizados, o
78 bien pueden conducir a realizar monitoreos durante un periodo menor al deseado (Soares et al.,
79 2023). Alternativamente, puede existir la necesidad de recopilar información existente antes o en
80 lugar de implementar protocolos de trabajo de campo; por ejemplo, los Estudios Previos
81 Prospectivos necesarios para el establecimiento de nuevas Áreas Naturales Protegidas en México
82 (DOF, 2014), regularmente incluyen listados de especies con presencia potencial en el área de
83 interés que son derivados de búsquedas bibliográficas (e.g. CONANP, 2014, 2023; SEDEMA,
84 2023a). Similarmente, diversas guías e instructivos oficiales para la elaboración de estudios
85 ambientales en México sugieren determinar la presencia potencial de especies en el área de
86 interés sin especificar un esfuerzo de muestreo mínimo (e.g. SEDEMA, 2023b; SEMARNAT,
87 2002, 2017), lo cual puede llevar a la realización de monitoreos limitados que son
88 complementados con información bibliográfica (e.g. CAM, 2009). Estos casos y otros similares
89 implican la realización de búsquedas bibliográficas o de trabajo de campo restringido temporal
90 y/o espacialmente (e.g. pocos muestreos durante un periodo limitado, o muestreos en una porción
91 limitada del área de interés), teniendo como consecuencia una caracterización biótica incompleta
92 o parcial.

93 Una alternativa para complementar estos métodos es el uso de datos abiertos disponibles
94 en plataformas digitales de biodiversidad. Gracias al esfuerzo de instituciones como la Comisión
95 Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO) en México y otras
96 similares alrededor del mundo (Urquiza-Haas et al., 2016), actualmente contamos con sistemas
97 de información sobre biodiversidad que recopilan, curan, concentran y ponen a disposición del
98 público información geográfica sobre la ocurrencia de especies, dando pie a la informática de la
99 biodiversidad (Soberón y Peterson, 2004). Adicionalmente, en las últimas décadas han surgido

100 diversas iniciativas de ciencia ciudadana o comunitaria a través de las cuales miembros del
101 público general registran sus observaciones de vida silvestre en páginas de internet o
102 aplicaciones para dispositivos móviles (Urquiza-Haas, 2016). Por ejemplo, la CONABIO
103 coordina la plataforma *naturalista* (<https://www.naturalista.mx/>; CONABIO, 2020) donde
104 usuarios pueden registrar observaciones y registros fotográficos de animales o plantas y que es
105 parte de la red *iNaturalist* (<https://www.inaturalist.org/sites/network>), así como *aVerAves*
106 (<https://ebird.org/averaves/home>; CONABIO, 2020), donde observadores de aves registran sus
107 avistamientos alimentando la plataforma de cobertura global *eBird* (<https://ebird.org/home>).
108 Estas y otras iniciativas de ciencia ciudadana, así como agencias gubernamentales e instituciones
109 educativas y de investigación de múltiples países, comparten su información sobre biodiversidad
110 con la *Global Biodiversity Information Facility* (GBIF).

111 GBIF sólo recibe datos de instituciones participantes que han suscrito el compromiso de
112 validar, curar, y organizar su información (<https://www.gbif.org/publishing-data>), además de
113 ponerla en un formato estándar diseñado para promover la compatibilidad entre datos
114 provenientes de distintas fuentes (Wieczorek et al., 2012). El rigor del proceso para contribuir
115 datos a GBIF hacen que sea una fuente confiable de información, aunque ésta siempre debe ser
116 validada por personas expertas (Maldonado et al., 2015). Por otro lado, la cantidad de
117 instituciones que contribuyen a su acervo, así como su distribución alrededor del mundo (2,441
118 publicadores al 12 de febrero de 2024; <https://www.gbif.org/the-gbif-network>), hacen de GBIF la
119 mayor colección de registros sobre biodiversidad a nivel global almacenados bajo un formato
120 común (Chandler et al., 2017; Soberón, 2022). Sin embargo, al recibir información de diversas
121 fuentes como museos, colecciones biológicas y datos históricos, GBIF guarda registros de
122 especies fósiles, especies que ya no se encuentran en el área donde fueron originalmente
123 registradas, o simplemente que tienen su nomenclatura taxonómica desactualizada (Maldonado
124 et al. 2015, Zizka et al., 2019, Jin y Yang, 2020). Por tal motivo es necesario realizar una
125 curación de los registros descargados de dicha plataforma (Zizka et al., 2020). Contar con un
126 método estandarizado y semiautomatizado que permita el acceso a datos de GBIF facilitaría la
127 recopilación y uso de información de biodiversidad.

128 En este trabajo nos preguntamos si la información disponible en GBIF puede
129 complementar inventarios biológicos realizados a través de monitoreos y muestreos en campo.
130 Específicamente, nuestro objetivo fue analizar si datos de ocurrencia de especies descargados de
131 dicha plataforma complementan listados de especies creados con trabajo de campo y reportados
132 en la literatura, utilizando a las Aves como caso de estudio. Para responder esta pregunta,
133 comparamos la complementariedad de listados de especies de aves publicados en artículos
134 arbitrados con aquellos generados con información disponible en GBIF para las mismas áreas de
135 estudio.

136 Con al menos 1,108 especies de aves, México ocupa el undécimo lugar entre los países
137 con mayor diversidad de esta clase taxonómica (Berlanga et al., 2015, Navarro et al. 2014).
138 Además de su riqueza, la facilidad de detección de diversas especies de aves contribuyen a la
139 realización de inventarios formales de este grupo taxonómico; por ejemplo, la participación de
140 *eBird* ha hecho de las aves el grupo más representado en GBIF debido a su popularidad entre la
141 sociedad (Troudet et al., 2017). Por tanto, consideramos que evaluar si los datos de GBIF
142 complementan listados de aves publicados, puede ilustrar el potencial de dicha plataforma para
143 complementar inventarios de taxones menos conspicuos en campo. Por ello, proporcionamos
144 como material suplementario el código de R desarrollado para 1) obtener registros de ocurrencias
145 de especies para cualquier área de interés, accediendo a GBIF, 2) limpiar dichos registros de

146 posibles errores geográficos y de nomenclatura, y 3) generar listados de especies. Este código
147 puede ser adaptado fácilmente para descargar registros de ocurrencias de cualquier lugar y grupo
148 taxonómico.

149

150

MATERIALES Y MÉTODOS

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

Realizamos una búsqueda de literatura dirigida para localizar artículos arbitrados que proporcionaran listados de especies de aves generados para áreas de estudio al interior de México. De cada artículo extrajimos la lista de especies reportada por los autores. En casos donde los listados fueron complementados a través de búsquedas de literatura, nos aseguramos de tomar únicamente los registros de especies identificadas a través de trabajo de campo. Además, extrajimos las coordenadas de la o las áreas de estudio de cada artículo; en casos en que las coordenadas no fueron reportadas o que la ubicación resultante no coincidiera con la mostrada en las figuras del área de estudio de los artículos, estimamos las coordenadas usando como guía dichas figuras. Utilizamos las coordenadas extraídas de los artículos, o las estimadas con sus figuras de estudio, para generar polígonos dentro de los cuales buscar ocurrencias de especies en GBIF. La cantidad de pares de coordenadas disponibles o estimadas de los artículos varió entre 1 y 6 (i.e. un sólo punto y un polígono irregular de 6 aristas). Para el caso de áreas de estudio identificadas por un sólo punto y dado que es necesario contar con un área definida para obtener las ocurrencias de especies registradas en su interior, generamos áreas (*buffers*) cuasi-circulares con un radio de 2.5 km. Para la realización de monitoreos de aves se recomienda la implementación de puntos de conteo separados entre 75 – 150 m o 250 m dependiendo de si son intensivos o extensivos, o de transectos que deben ser recorridos a una velocidad constante avanzando 100 m o 250 m cada 10 minutos (Ralph et al., 1996). Consideramos que un área circular con radio de 2.5 km es suficiente para establecer un monitoreo de aves tradicional que siga esas prácticas. Para el caso de áreas de estudio definidas por dos pares de coordenadas generamos el polígono rectangular mínimo alrededor del área de estudio (*bounding box*), ya que este polígono mínimo abarca por definición el área de estudio real. Para áreas de estudio con >2 pares de coordenadas, generamos el polígono correspondiente para emular de manera más cercana el área donde los autores de esos estudios realizaron sus observaciones. Tanto los polígonos mínimos como aquellos generados para estudios que proporcionaron más de dos pares de coordenadas, son representaciones cercanas de las áreas de estudio reales reportadas en los artículos consultados. Por tanto, decidimos no agregar *buffers* a esos casos, para minimizar el riesgo de descargar de GBIF registros de especies realizados fuera de las áreas de estudio reportadas en los artículos.

Una vez replicadas las áreas de estudio, utilizamos la función ‘occ_download’ del paquete ‘rgbif’ (Chamberlain et al., 2022) en el lenguaje de programación R versión 4.3.0 (R Core Team, 2023), para descargar de GBIF registros de aves al interior del polígono de cada área de estudio. Para esto proporcionamos a la función las coordenadas de cada polígono en formato WKT y especificamos que los registros a buscar 1) deberían ser de Aves, 2) deberían estar georeferenciados, 3) su información geoespacial no debería tener problemas, 4) deberían haber sido realizados después del año 1900, 4) deberían ser de individuos presentes al interior del polígono y 5) no deberían incluir registros fósiles. En el material suplementario se muestra la estructura del código utilizado y la manera en que estos argumentos fueron especificados.

La calidad de los datos disponibles en GBIF varía de acuerdo su fuente de origen en tres dimensiones principales: temporal, espacial, y taxonómica (Jin y Yang, 2020). Por ejemplo, los datos provenientes de una colección biológica depositada en algún museo pueden tener fechas de

192 recolección incorrectas, errores en las coordenadas geográficas, nombres científicos
193 desactualizados, entre otros. Además, debido a que GBIF concentra información de diversas
194 fuentes, es común que existan registros de especies duplicados para una misma región. Por este
195 motivo, revisamos y limpiamos los registros descargados de GBIF utilizando el paquete de R
196 'bdc' (Ribeiro et al., 2022a) siguiendo el protocolo descrito por Ribeiro et al. (2022b).

197 En resumen, el protocolo de limpieza de datos descargados de GBIF se divide en 4
198 grandes pasos (Ribeiro et al., 2022b). En primer lugar, se realiza un prefiltrado para detectar y
199 eliminar registros sin nombre científico, sin coordenadas o con coordenadas erróneas, o de
200 naturaleza que no sea de interés para un listado de especies actuales en el área de estudio (e.g.
201 fósiles), entre otros. En segundo lugar, se realiza una revisión taxonómica para asegurarse de que
202 todos los registros tengan un nombre científico actualizado. Este paso identifica registros cuyos
203 nombres científicos no pudieron ser identificados o actualizados por 'bdc'; en caso de ocurrir,
204 este tipo de registros los revisamos y cotejamos manualmente con los nombres reconocidos por
205 el listado de aves de Norte y Centro América hasta su 63° suplemento (Chesser et al., 2022) y en
206 caso de ser reconocidos por esa autoridad taxonómica los mantuvimos en los listados. Una
207 consideración importante al usar datos de GBIF y otras bases de datos de biodiversidad es que
208 las coordenadas geográficas de los registros de especies pueden ser inexactas (Maldonado et al.,
209 2015). Por ejemplo, registros antiguos pudieron ser referenciados con el nombre de la localidad
210 de observación y no con coordenadas, o bien no existe mayor referencia que el país de registro;
211 el problema surge en que al ser digitalizados, este tipo de registros pueden recibir las
212 coordenadas del centroide geográfico del país o localidad de registro, o de la institución que
213 alberga o digitalizó la observación (Zizka et al., 2019). Por tanto, el tercer paso consiste en una
214 validación espacial de los registros; por ejemplo, que sus coordenadas no correspondan con la
215 ubicación de museos, colecciones, o alguna otra institución reconocida. El último paso consiste
216 en una validación temporal que identifica registros sin fecha o año de observación; además, es
217 posible especificar la antigüedad de los registros a considerar (e.g. eliminar registros previos a
218 cierto año). Después de aplicar este protocolo a los datos descargados para cada área de estudio,
219 eliminamos registros duplicados, generando así un listado de especies únicas reportadas en cada
220 área.

221 Desde el año 2002, la nomenclatura taxonómica de las Aves de Norte y Centro América
222 es revisada y actualizada anualmente ([https://americanornithology.org/publications/north-and-
223 middle-american-checklist/](https://americanornithology.org/publications/north-and-middle-american-checklist/)). Debido a que los artículos encontrados en nuestra búsqueda fueron
224 publicados a lo largo de 3 décadas (1991 – 2022), sus listados fueron elaborados siguiendo
225 normas taxonómicas distintas. Por tanto, a los listados disponibles en artículos les aplicamos el
226 mismo protocolo de revisión y actualización taxonómica descrito arriba.

227 La actualización de nomenclatura taxonómica de los datos descargados de GBIF
228 identificó un total de 425 registros de 334 subespecies, resaltando el potencial de la plataforma
229 para proporcionar registros a nivel infraespecífico. En comparación, el mismo método de
230 actualización sugirió tan sólo tres subespecies de entre todas las especies reportadas en los
231 artículos consultados. Para comparar listados usando el mismo nivel taxonómico, eliminamos de
232 los registros de GBIF aquellas subespecies cuyas especies estuvieran incluidas en el mismo
233 listado (e.g. en caso de que un listado generado con datos de GBIF estuviera reportada
234 *Sporophila torqueola* y *S. t. morelleti*, sólo dejamos la primera). Además, revisamos si alguna de
235 las subespecies en cada listado generado con GBIF estaba reportada a nivel de especie en los
236 listados de los artículos correspondientes; en caso de ser así, para que la comparación pudiera
237 realizarse a nivel de especie, eliminamos del registro del GBIF el epíteto infraespecífico.

238 Finalmente, comparamos los listados de especies reportados en cada publicación con los
239 descargados de GBIF con el índice de disimilitud de Sorensen (β_{sor}) y sus dos componentes
240 (recambio β_{sim} y anidamiento β_{sne}) utilizando la función 'beta.pair' del paquete de R 'betapart'
241 (Baselga et al., 2022). Elegimos el índice de Sorensen porque tanto éste mismo como sus
242 componentes responden de manera lineal al traslape gradual en la composición de especies de las
243 muestras comparadas, facilitando su interpretación (MacGregor-Fors et al., 2022).

244 245 **Resultados**

246 Encontramos 22 artículos arbitrados publicados entre 1991 y 2022 que reportan listados
247 de aves obtenidos a través de trabajo de campo en México (Tabla 1); uno de ellos reportó
248 coordenadas y listados para tres localidades distintas (Mellink, 1991), por lo que obtuvimos un
249 total de 24 listados para igual número de áreas de estudio. Después de limpiar y depurar los
250 listados publicados, el número mínimo y máximo de especies reportadas como registros en
251 campo varió entre 45 y 245. GBIF no contó con registros de aves para tres de las 24 áreas de
252 estudio; por tanto, el menor número de especies disponibles en esa plataforma para las áreas de
253 estudio fue de 0, pero el mayor fue de 435 (Tabla 1).

254
255
256
257

TABLA 1 / TABLE 1

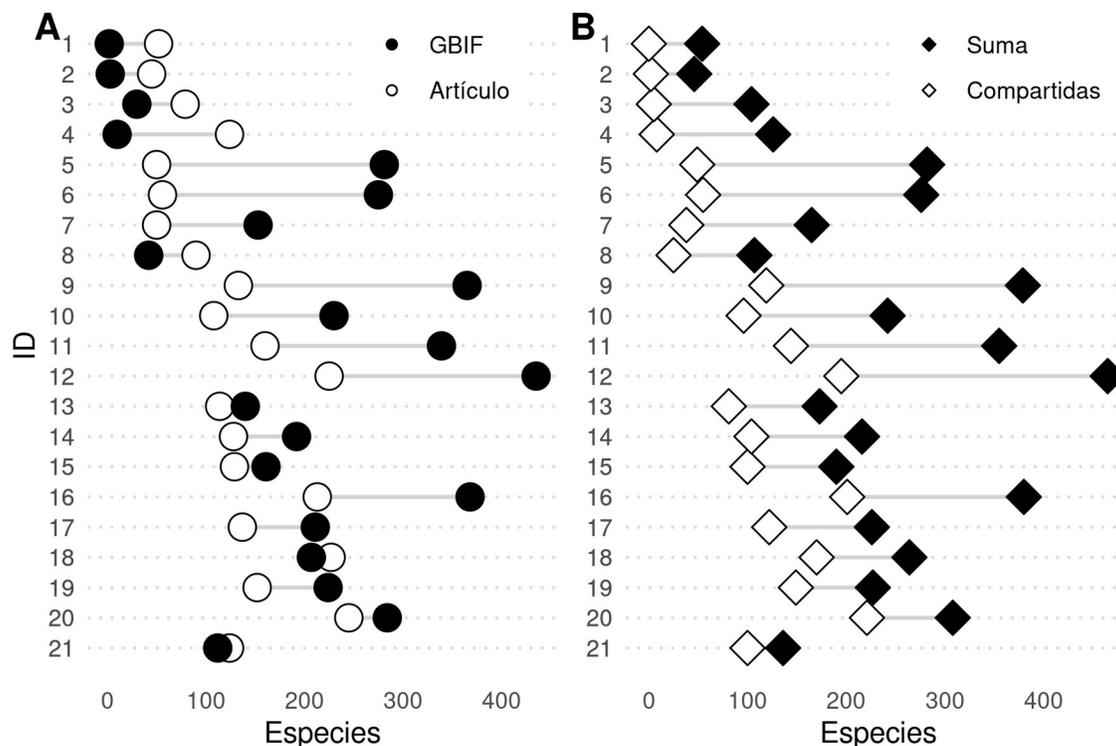
Comparación de la riqueza de especies entre listados obtenidos de dos fuentes de información para las mismas áreas de estudio, e índices de disimilitud en la composición de especies. / Comparison of species richness between lists obtained from two sources of information for the same study areas, and dissimilarity indices in the composition of species.

ID*	Spp. Artículo		Spp. GBIF		Especies			%↑	Disimilitud			GBIF DOI
	Total	Únicas	Total	Únicas	Compartidas	Suma	Diferencia		β_{sor}	β_{sim}	β_{sne}	
1	52	52	2	2	0	54	50	3.8	1	1	0	doi.org/10.15468/dl.vumns4
2	45	43	3	1	2	46	42	2.2	0.92	0.33	0.58	doi.org/10.15468/dl.jdze7e
3	79	74	30	25	5	104	49	31.6	0.91	0.83	0.07	doi.org/10.15468/dl.5p6qtd
4	124	116	10	2	8	126	114	1.6	0.88	0.2	0.68	doi.org/10.15468/dl.x5mp7d
5	50	1	281	232	49	282	-231	464.0	0.7	0.02	0.68	doi.org/10.15468/dl.8wev58
6	56	1	275	220	55	276	-219	392.9	0.67	0.02	0.65	doi.org/10.15468/dl.b4zx6z
7	50	12	153	115	38	165	-103	230.0	0.63	0.24	0.39	doi.org/10.15468/dl.sbezde
8	90	65	42	17	25	107	48	18.9	0.62	0.4	0.22	doi.org/10.15468/dl.jawt3k
9	133	14	365	246	119	379	-232	185	0.52	0.11	0.42	doi.org/10.15468/dl.523w8z
10	108	12	230	134	96	242	-122	124.1	0.43	0.11	0.32	doi.org/10.15468/dl.ev9km8
11	160	16	339	195	144	355	-179	121.9	0.42	0.1	0.32	doi.org/10.15468/dl.r7ppcb
12	225	30	435	240	195	465	-210	106.7	0.41	0.13	0.28	doi.org/10.15468/dl.fvcun6
13	114	33	140	59	81	173	-26	51.8	0.36	0.29	0.07	doi.org/10.15468/dl.mu8xu3
14	128	24	192	88	104	216	-64	68.8	0.35	0.19	0.16	doi.org/10.15468/dl.qmyv7y
15	213	12	368	167	201	380	-155	78.4	0.31	0.06	0.25	doi.org/10.15468/dl.23ynr6
16	129	29	161	61	100	190	-32	47.3	0.31	0.22	0.09	doi.org/10.15468/dl.xbnce2
17	137	15	211	89	122	226	-74	65	0.3	0.11	0.19	doi.org/10.15468/dl.axwr94
18	227	57	207	37	170	264	20	16.3	0.22	0.18	0.04	doi.org/10.15468/dl.74auta
19	152	3	224	75	149	227	-72	49.3	0.21	0.02	0.19	doi.org/10.15468/dl.usmxk3
20	245	24	284	63	221	308	-39	25.7	0.16	0.1	0.07	doi.org/10.15468/dl.fkw4zw
21	124	24	112	12	100	136	12	9.7	0.15	0.11	0.05	doi.org/10.15468/dl.df63bj
22	61	61	0	0	0	61	-61	0	-	-	-	doi.org/10.15468/dl.3gbpst
23	56	56	0	0	0	56	-56	0	-	-	-	doi.org/10.15468/dl.4w3yth
24	93	93	0	0	0	93	-93	0	-	-	-	doi.org/10.15468/dl.gdbmhd

258 ID: ¹Mellink (1991) localidad Las Joyas, ²Arellano-Delgado et al. (2018), ³Castillo-Guerrero et al. (2009), ⁴Rojas-Soto et al. (2009),
 259 ⁵MacGregor-Fors, Gómez-Martínez, et al. (2022), ⁶MacGregor-Fors y Schondube (2011), ⁷Ugalde-Lezama et al. (2010), ⁸Martínez-
 260 Morales et al. (2013), ⁹Ramírez-Albores (2006), ¹⁰Ruiz-Campos y Rodríguez-Meraz (1997), ¹¹Ramírez-Albores et al. (2018),
 261 ¹²Ramírez-Albores (2010), ¹³Bojorges-Baños y López-Mata (2006), ¹⁴Ramírez-Albores (2007), ¹⁵Santiago-Alarcon et al. (2011),
 262 ¹⁶Ramírez-Albores (2013), ¹⁷Canales-Delgadillo et al. (2020), ¹⁸Manzano-Fischer et al. (2006), ¹⁹Castillo-Muñoz y Guzmán-
 263 Hernández (2021), ²⁰Sahagún-Sánchez et al. (2021), ²¹Adame et al. (2019), ²²Mellink (1991) localidad Las Canteras, ²³Mellink (1991)

266 localidad San Antonio, ²⁴Gómez-Moreno et al. (2022). Leyendas: Spp. Artículo = Número de especies publicadas en artículos; Spp.
267 GBIF = Número de especies registradas en GBIF. Total = Número total de especies en los listados; Únicas = número de especies no
268 compartidas entre fuentes. Compartidas = número de especies reportadas en ambas fuentes; Suma = Número de especies únicas
269 reportadas entre ambas fuentes; Diferencia = Número total de especies reportadas en artículos menos el total reportado en GBIF, por lo
270 que números negativos indican mayor número de especies en GBIF. %↑ = Porcentaje de incremento entre las especies reportadas en
271 los artículos (columna Spp. Artículo | Total) y las especies únicas reportadas entre ambas fuentes (columna Suma). β_{sor} = Índice de
272 disimilitud de Sorensen, β_{sim} y β_{sne} = Recambio y anidamiento de especies respectivamente. La tabla está ordenada de acuerdo al valor
273 decreciente de β_{sor} . La columna GBIF DOI contiene el enlace al sitio desde donde se pueden volver a descargar los datos de cada área
274 de estudio. Legend: Spp. Artículo = Number of species in the published papers; Spp. GBIF = Number of species documented in GBIF.
275 Total = total number of species listed; Únicas (unique) = number of species not shared between sources; Compartidas (shared) =
276 number of species reported by both sources; Suma (sum) = Number of unique species reported between both sources; Diferencia
277 (difference) = Total number of species reported in papers minus total number of species reported in GBIF, so that negative values
278 indicate higher species number in GBIF. %↑ = Increase percent between species listed in articles (column Spp. Artículo | Total) and
279 unique species reported between both sources (column Sum). β_{sor} = Sorensen dissimilarity index, β_{sim} and β_{sne} = Species turnover and
280 species nestedness respectively. This table is ordered according to the decreasing value of β_{sor} . The GBIF DOI column contains a link
281 to the site where the occurrence of species data can be downloaded again for each study area.

282 GBIF documentó la presencia de entre 26 – 232 especies más que las reportadas en artículos
 283 para 14 de las áreas de estudio (Tabla 1 y Fig. 1A), evidenciando el potencial de la plataforma para
 284 complementar inventarios avifaunísticos. Por otro lado, los listados publicados en artículos para 10 de
 285 las áreas de estudio contuvieron entre 12 – 114 especies más que las registradas en GBIF (Tabla 1 y
 286 Fig. 1A), resaltando el valor de los muestreos en campo para continuar documentando la riqueza
 287 biológica de áreas poco exploradas.
 288



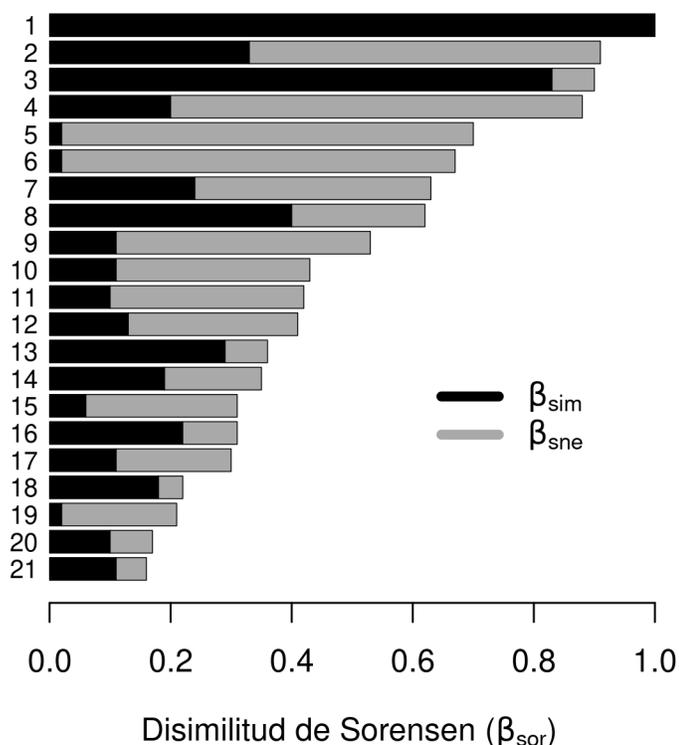
289 **Fig. 1. A.** Número de especies listadas en artículos y en GBIF para cada área de estudio para las cuáles
 290 se pudieron realizar comparaciones. **B.** Número de especies reportadas en ambas fuentes (compartidas)
 291 y suma de especies únicas reportadas entre ambas fuentes, ilustrando la riqueza adicional obtenida al
 292 combinar información entre fuentes. El eje Y se corresponde con el código ID en la Tabla 1.
 293

294 **Fig. 1.** Number of species listed in papers and in GBIF for each study area for which comparisons
 295 could be made. **B.** Number of species reported in both sources (shared) and sum of unique species
 296 reported between both sources, illustrating the additional richness obtained when combining
 297 information from both sources. Values in the Y axis corresponds with the ID code in Table 1.
 298

299 El número de especies registradas exclusivamente en los listados de los artículos varió entre 1 y
 300 116, mientras que el número de especies registradas exclusivamente en GBIF, varió entre 0 y 246
 301 (Tabla 1). Puesto de otra manera, los datos de GBIF aportaron hasta 246 especies a los listados
 302 reportados en los artículos consultados, mientras que las especies reportadas en publicaciones aportaron
 303 hasta 116 especies adicionales a las documentadas en GBIF (Fig. 1B). La cantidad de especies obtenida
 304 al combinar información de los artículos y de GBIF incrementó entre 1.6% y hasta 464% comparada
 305 con la listada solamente en los artículos (Tabla 1).

306 El índice de disimilitud de Sorensen (β_{sor}) para las 21 áreas de estudio para las que GBIF tuvo
 307 especies registradas varió entre 0.15 y 1 (Tabla 1), mostrando que siempre hubo diferencias en la
 308 composición de especies entre listados. El valor menor de β_{sor} (= 0.15) se desprendió de la comparación
 309 entre los listados con la menor diferencia en su número de especies (124 en el artículo, 112 en GBIF;
 310 Tabla 1, ID 21); a pesar del similar número de especies reportadas para esa área en particular, la
 311 disimilitud estuvo explicada principalmente por recambio de especies (Fig. 2), mostrando una vez más

312 la complementariedad entre fuentes de información. Por otro lado, el valor mayor de β_{sor} ($= 1$) indica
 313 que GBIF documentó para un área especies completamente distintas a las reportadas en el artículo.
 314 Además, β_{sim} (recambio) nunca fue 0 indicando que una fuente de información siempre listó especies
 315 que no estuvieron reportadas en la otra. De manera similar, β_{sne} nunca alcanzó un valor de 1, indicando
 316 que el anidamiento de especies entre fuentes de información nunca fue total.
 317



318 **Fig. 2.** Disimilitud total en la composición de especies (β_{sor} , barras enteras) entre listados publicados y
 319 aquellos generados con datos de GBIF. La porción negra y gris de cada barra representa la disimilitud
 320 por recambio de especies (β_{sim}) y por anidamiento (β_{sne}) respectivamente. Cada barra corresponde a un
 321 área de estudio, el eje Y hacen referencia al código ID en la Tabla 1.
 322

323 **Fig. 2.** Total dissimilarity in species composition (β_{sor} , full bars) between published lists and those
 324 generated with GBIF data. The black and gray portions of each bar represent dissimilarity by species
 325 turnover (β_{sim}) and by nestedness (β_{sne}) respectively. Each bar corresponds to a study area, values in the
 326 Y axis refer to the ID code in Table 1.
 327

328

329

DISCUSIÓN

330

331 La cantidad de especies reportadas para cada una de las áreas de estudio fue hasta 464% mayor
 332 al combinar la información disponible en los artículos y en GBIF, que al considerar los registros de una
 333 sola fuente de información. De hecho, los registros disponibles en GBIF contribuyeron con hasta 246
 334 especies a los listados reportados en los artículos. Por tanto, la información disponible en GBIF puede
 335 complementar con un número importante de especies la colectada a través de muestreos y monitoreos
 en campo.

336

337 Utilizar datos disponibles en GBIF para complementar esfuerzos de muestreo permite
 338 aprovechar el tiempo, esfuerzo y los recursos invertidos por instituciones nacionales y académicas para
 339 recolectar información. Por ejemplo, el costo promedio absorbido por CONABIO para adquirir datos
 340 primarios de biodiversidad es de aproximadamente \$5,500 dólares americanos por proyecto apoyado
 341 (Soberón, 2022) y el costo promedio por digitalizar esa información para hacerla disponible a través
 plataformas es de 8 dólares (Jiménez et al., 2016). GBIF también reúne datos recolectados por personas

342 físicas. Actualmente, los registros de biodiversidad provenientes de plataformas de ciencia ciudadana
343 representan al menos 60% de las observaciones disponibles en GBIF (Groom et al., 2017), rebasando la
344 cantidad de información generada a través de esfuerzos académicos formales. Por tal motivo, la
345 información concentrada en GBIF es resultado de la suma de inversiones hechas por múltiples actores,
346 por lo que puede contener registros de especies raras o poco comunes que monitoreos formales
347 realizados por profesionales podrían no registrar (e.g. Zhou et al., 2020; cit. por Ding et al., 2022).

348 Si bien se ha cuestionado la validez o el rigor de observaciones hechas por científicos
349 ciudadanos (Aceves-Bueno et al., 2017), los observadores de aves que contribuyen a plataformas
350 digitales tienen un grado de especialización en sus habilidades de identificación superior al de personas
351 que no contribuyen (Randler, 2021). De hecho, los participantes más activos suelen tener un grado alto
352 de especialización (Rosenblatt et al., 2022). Por tanto, se puede considerar que los registros de aves
353 contenidos en esas plataformas son confiables. Sin embargo, plataformas como *iNaturalist* introducen
354 incertidumbre a las coordenadas geográficas de registros de especies amenazadas con la finalidad de
355 protegerlas, limitando el uso de esos datos (Contreras-Díaz et al., 2023). Por tanto, los usuarios deben
356 tomar medidas necesarias para incrementar la confianza en los datos descargados desde plataformas
357 digitales, por ejemplo teniendo conocimiento de sus políticas de manejo de datos y aplicando los filtros
358 necesarios a la información obtenida.

359 El potencial de los datos en GBIF para complementar esfuerzos de campo variará por grupo
360 taxonómico. Por ejemplo, las aves están hiper-representadas en GBIF, mientras que muchos otros
361 grupos taxonómicos como los insectos están severamente sub-representados (Troudet et al., 2017); por
362 tanto, la cantidad de información disponible para complementar esfuerzos de campo es distinta para
363 cada taxón. Por otro lado, para tener una certidumbre alta en la identificación de algunos grupos
364 taxonómicos hecha por científicos ciudadanos a través de registros fotográficos, es necesario el
365 consenso de al menos 5 personas (Swanson et al., 2016). Sin embargo, la plataforma *iNaturalist*
366 requiere del consenso de tan sólo 2 – 3 personas para otorgar el “grado académico” a las
367 identificaciones hechas por sus usuarios (<https://www.inaturalist.org/pages/help#quality>), grado con el
368 cual la observación pasa a ser reportada a GBIF ([https://www.gbif.org/dataset/50c9509d-22c7-4a22-](https://www.gbif.org/dataset/50c9509d-22c7-4a22-a47d-8c48425ef4a7#description)
369 [a47d-8c48425ef4a7#description](https://www.gbif.org/dataset/50c9509d-22c7-4a22-a47d-8c48425ef4a7#description)). Posiblemente debido a esto, la identidad de algunos anfibios y
370 reptiles registrados en GBIF a través de *iNaturalist* es errónea (JLAL, observación personal). Salvo por
371 grupos de especies con apariencia muy parecida entre sí, la identificación de aves en plataformas de
372 ciencia ciudadana es generalmente correcta (Gorleri et al., 2023); por tanto, los registros fauna no aviar
373 disponibles en GBIF podrían requerir una validación más cuidadosa, especialmente si provienen de
374 plataformas de ciencia ciudadana.

375 Los datos disponibles en GBIF son útiles e informativos para ciertas tareas, pero no sustituyen
376 los monitoreos formales realizados por profesionales (Danielsen et al., 2005; Kamp et al., 2016). De
377 hecho, nuestros resultados muestran que la complementación de listados es bidireccional: los listados
378 contenidos en los artículos enriquecieron con al menos una y hasta 116 especies los generados con
379 registros de GBIF. Si bien GBIF constituye el mayor repositorio de datos de biodiversidad a nivel
380 global, puede continuar siendo alimentado. A pesar de que las aves son uno de los grupos con más
381 registros en GBIF, algunas de las áreas de estudio consultadas tienen pocos registros o ninguno,
382 ilustrando que quedan zonas que no han sido suficientemente muestreadas (Hughes et al., 2021), o que
383 la información existente no ha llegado a la plataforma. Por ejemplo, 10 de los artículos consultados
384 listan especies no documentadas en GBIF para sus áreas de estudio por lo que su información podría
385 ser un gran aporte a la plataforma. De manera similar, observaciones de biodiversidad reportadas en
386 Facebook, que no fue diseñado para recopilar ese tipo de información, contienen registros que
387 complementan la información disponible en GBIF (Chowdhury et al., 2023). Por tal motivo, queremos
388 resaltar la importancia de continuar realizando muestreos de campo cuando la situación lo requiera y
389 los recursos lo permitan, haciendo simultáneamente un llamado a incorporar la información colectada a
390 plataformas de datos abiertos, especialmente la pertenecientes a regiones o grupos taxonómicos
391 subrepresentados (Troudet et al., 2017), para contribuir a subsanar los sesgos geográficos y

392 taxonómicos de la información disponible en GBIF y para que la información arduamente generada
 393 enriquezca el acervo de datos abiertos y sea accesible a todo el público.

394 Los datos que se contribuyen a GBIF no necesitan contener grandes números de registros o de
 395 especies. Por ejemplo, el juego de datos “*Biodiversidad selecta de los humedales de Laguna de*
 396 *Términos - Pantanos de Centla (Aves)*” ([https://www.gbif.org/es/dataset/9b5720a2-99f7-4e29-86c8-](https://www.gbif.org/es/dataset/9b5720a2-99f7-4e29-86c8-c028e67a86f2)
 397 [c028e67a86f2](https://www.gbif.org/es/dataset/9b5720a2-99f7-4e29-86c8-c028e67a86f2)) reporta únicamente dos especies de aves recolectadas y preservadas en 1998 (Arriaga
 398 Weiss, 2021). Estos datos han sido descargados más de 6000 veces y tienen 30 citas. De manera
 399 similar, los datos en GBIF tampoco necesitan provenir de esfuerzos de muestreo prolongados. Por
 400 ejemplo, el juego de datos “*Mycoblitz 2017 - IRBV et Cercle des Mycologues de Montréal*”
 401 (<https://www.gbif.org/es/dataset/39bd4817-9530-4341-a918-4f16b8db1419>) consiste en 134
 402 ocurrencias de macro hongos colectados durante sólo 3 horas de un sólo día en el *Parc des chutes de*
 403 *Saint-Ursule* en Quebec, Canadá (Archambault y Sinou, 2017). Este juego de datos ha sido descargado
 404 más de 6000 veces y ha sido citado en 26 ocasiones.

405 Nuestros resultados demuestran una vez más que combinar datos disponibles en GBIF con los
 406 de otras fuentes resulta en inventarios biológicos más completos (de Araujo et al., 2022). Nuestro
 407 trabajo no tuvo como objetivo comparar cuál fuente (GBIF o artículos) proporciona inventarios más o
 408 menos completos, sino analizar el potencial que los datos en GBIF tienen para complementar
 409 inventarios publicados. Por tanto, el uso del año 1900 como umbral a partir del cual descargar registros,
 410 tuvo la intención de aprovechar la mayor cantidad de información acumulada en GBIF para cada área
 411 de estudio.

412 El proceso de generación de listados de especies utilizando registros de GBIF con el código
 413 proporcionado en el material suplementario toma sólo unos minutos, por lo que este trabajo no sólo
 414 ilustra la posibilidad de complementar inventarios biológicos utilizando datos abiertos, sino que facilita
 415 el acceso y recopilación de información sobre biodiversidad que ecólogos, estudiantes y prestadores de
 416 servicios profesionales en materia ambiental realizan rutinariamente, reduciendo el tiempo necesario
 417 para esta tarea. Sin embargo, los listados generados usando datos de GBIF deben ser revisados
 418 cuidadosamente para evitar la comisión de errores (Maldonado et al., 2015), e idealmente no deberían
 419 usarse en sustitución de monitoreos de campo, sino como complemento de inventarios derivados de
 420 trabajo de campo formal.

421

422 AGRADDECIMIENTOS

423 Agradecemos a Tereso Hernández Morales por su apoyo en la desambiguación de nombres científicos
 424 no resueltos por ‘bdc’, así como a Rafael Rueda Hernández por probar el código disponible en el
 425 material suplementario y por sus comentarios sobre el mismo. Gracias a cuatro revisores anónimos
 426 cuyos comentarios ayudaron a mejorar este manuscrito.

427

428 REFERENCIAS

Aceves-Bueno, E., Adeleye, A. S., Feraud, M., Huang, Y., Tao, M., Yang, Y., & Anderson, S. E. (2017).
 The accuracy of citizen science data: A quantitative review. *The Bulletin of the Ecological Society of*
America, 98(4), 278–290. <https://doi.org/10.1002/bes2.1336>

Adame, D. R., Moranchel, A. L., & Piedragil, C. D. J. (2019). Avifauna del parque estatal Cerro de la
 Tortuga, Morelos, México. *Mesoamericana*, 23(1), Article 1.

Archambault, R., & Sinou, C. (2017). *Mycoblitz 2017—IRBV et Cercle des Mycologues de Montréal.*
Université de Montréal Biodiversity Centre. Occurrence dataset <https://doi.org/10.5886/ebybuj>
 accessed via GBIF.org on 2023-03-13. <https://doi.org/10.5886/ebybuj>

Arellano-Delgado, M., Guerra-Quesada, F., Cabrera-Manuel, F., Sandoval-Ortega, M. H., & Araiza-Arvilla, H. N. (2018). Avifauna de la comunidad de La Lumbreira (Cieneguilla), Aguascalientes, Aguascalientes, México. *Tecnociencia Chihuahua*, 12(2), 114–123.

Arriaga Weiss, S. L. (2021). *Biodiversidad selecta de los humedales de Laguna de Términos—Pantanos de Centla (Aves). Versión 1.5. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad. Occurrence dataset <https://doi.org/10.15468/qgohxy> accessed via GBIF.org on 2023-03-13. <https://doi.org/10.15468/qgohxy>*

Baselga, A., Orme, D., Villegier, S., Bortoli, J. D., Leprieur, F., Logez, M., Martínez-Santalla, S., Martín-Devasa, R., Gómez-Rodríguez, C., Crujeiras, R., & Henriques-Silva, R. (2022). *betapart: Partitioning Beta Diversity into Turnover and Nestedness Components (1.5.6)* [Computer software]. <https://CRAN.R-project.org/package=betapart>

Berlanga, H. A., Gómez de Silva, H., Vargas Canales, V. M., Rodríguez Contreras, V., Sánchez González, L. A., Ortega Álvarez, R., & Calderón-Parra, R. (2015). Aves de México: Lista actualizada de especies y nombres comunes 2015 (Primera Edición). Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO). México.

Bojorges-Baños, J. C., & López-Mata, L. (2006). Asociación de la riqueza y diversidad de especies de aves y estructura de la vegetación en una selva mediana subperennifolia en el centro de Veracruz, México. *Revista Mexicana de Biodiversidad*, 77(2), 235–249.

Cabrera-Cruz, S. A., Aguilar López, J. L., Aguilar-Rodríguez, P. A., Oropeza-Sánchez, M. T., Muñoz Jiménez, O., & Villegas Patraca, R. (2023). Changes in diversity and species composition in the assemblage of live and dead bats at wind farms in a highly diverse region. *Environmental Monitoring and Assessment*, 195(12), 1480. <https://doi.org/10.1007/s10661-023-12090-z>

CAM. (2009). (Corporación Ambiental de México, S.A. de C.V). Manifestación de Impacto Ambiental, Modalidad Regional Energía Sierra Juárez, S. de R.L. de C.V. (MIA 09011; p. 453). Corporación Ambiental de México, S.A. de C.V. <https://app.semarnat.gob.mx/consulta-tramite/#/resultado-busqueda>. Consultado el 08 de febrero de 2024.

Canales-Delgadillo, J. C., Benítez-Orduña, E., Pérez-Ceballos, R. Y., Zaldívar Jiménez, A., Gómez-Ponce, M., Cardoso-Mohedano, J. G., & Merino-Ibarra, M. (2020). Inter-annual diversity of birds in the shoreline of an island in the southern Gulf of Mexico. *Huitzil*, 21(1). <https://doi.org/10.28947/hrmo.2020.21.1.433>

Castillo-Guerrero, J. A., González-Medina, E., & González-Bernal, M. A. (2009). Patrones de presencia y abundancia de aves terrestres en la isla Saliaca, Sinaloa, México. *Revista Mexicana de Biodiversidad*, 80(1), 211–218.

Castillo-Muñoz, M., & Guzmán-Hernández, J. L. (2021). Composición y estructura de la comunidad de aves en un corredor ribereño urbano del área metropolitana de Monterrey, Nuevo León, México. *Huitzil Revista Mexicana de Ornitología*, 22(2), Article 2. <https://doi.org/10.28947/hrmo.2021.22.2.499>

Chamberlain, S., Oldoni, D., Barve, V., Desmet, P., Geffert, L., Mcglinn, D., Ram, K., rOpenSci (<https://ropensci.org/>), Waller [aut, J., & cre. (2022). *rgbif: Interface to the Global Biodiversity Information Facility API (3.7.2)* [Computer software]. <https://CRAN.R-project.org/package=rgbif>

Chandler, M., See, L., Copas, K., Bonde, A. M. Z., López, B. C., Danielsen, F., Legind, J. K., Masinde, S., Miller-Rushing, A. J., Newman, G., Rosemartin, A., & Turak, E. (2017). Contribution of citizen science towards international biodiversity monitoring. *Biological Conservation*, 213, 280–294. <https://doi.org/10.1016/j.biocon.2016.09.004>

Chesser, R. T., Billerman, S. M., Burns, K. J., Cicero, C., Dunn, J. L., Hernández-Baños, B. E., Jiménez, R. A., Kratter, A. W., Mason, N. A., Rasmussen, P. C., Remsen, J. V., Stotz, D. F., & Winker, K. (2022). *Check-list of North American Birds (online)*. American Ornithological Society. Check-List of North and Middle American Birds. <https://checklist.americanornithology.org/taxa/>

Chowdhury, S., Aich, U., Rokonzaman, M., Alam, S., Das, P., Siddika, A., Ahmed, S., Labi, M. M., Marco, M. D., Fuller, R. A., & Callaghan, C. T. (2023). Increasing biodiversity knowledge through social media: A case study from tropical Bangladesh. *BioScience*, biad042. <https://doi.org/10.1093/biosci/biad042>

Colwell, R. K., Mao, C. X., & Chang, J. (2004). Interpolating, extrapolating, and comparing incidence-based species accumulation curves. *Ecology*, 85(10), 2717–2727. <https://doi.org/10.1890/03-0557>

CONABIO. (2020). (*Comisión Nacional para el Conocimiento y Uso de la Biodiversidad*). *Informe de actividades 2017–2019* (p. 39). Comisión Nacional para el Conocimiento y Uso de la Biodiversidad. <https://bioteca.biodiversidad.gob.mx/janium-bin/detalle.pl?Id=20230602120042>

CONANP. (2014). (Comisión Nacional de Áreas Naturales Protegidas). Estudio Previo Justificativo para el establecimiento del Área Natural Protegida Área de Protección de Flora y Fauna “Monte Mojino” ubicada en el estado de Sinaloa, México (p. 295). CONANP, SEMARNAT, CONSELVA, CIAD. <https://biblioteca.semarnat.gob.mx/janium/Documentos/Ciga/Libros2013/CD003137a.pdf>. Consultado el 08 de febrero de 2024.

CONANP. (2023). (Comisión Nacional de Áreas Naturales Protegidas). Estudio Previo Justificativo para el establecimiento del Área Natural Protegida Reserva de la Biósfera Wanha’. Tabasco, México. (p. 232) [Estudio Previo Justificativo]. CONANP, SEMARNAT, Universidad Juárez Autónoma de Tabasco. <https://www.conanp.gob.mx/pdf/separata/EPJ-RB-Wanha.pdf>. Consultado el 08 de febrero de 2024.

Contreras-Díaz, R. G., Nori, J., Chiappa-Carrara, X., Peterson, A. T., Soberón, J., & Osorio-Olvera, L. (2023). Well-intentioned initiatives hinder understanding biodiversity conservation: Cloaked iNaturalist information for threatened species. *Biological Conservation*, 282, 110042. <https://doi.org/10.1016/j.biocon.2023.110042>

Danielsen, F., Burgess, N. D., & Balmford, A. (2005). Monitoring Matters: Examining the Potential of Locally-based Approaches. *Biodiversity & Conservation*, 14(11), 2507–2542. <https://doi.org/10.1007/s10531-005-8375-0>

de Araujo, M. L., Quaresma, A. C., & Ramos, F. N. (2022). GBIF information is not enough: National database improves the inventory completeness of Amazonian epiphytes. *Biodiversity and Conservation*. <https://doi.org/10.1007/s10531-022-02458-x>

Ding, Y., Xiong, L., Ji, F., Lu, J., Zhu, X., & Huang, H. (2022). Using citizen science data to improve regional bird species list: A case study in Shaanxi, China. *Avian Research*, 13, 100045. <https://doi.org/10.1016/j.avrs.2022.100045>

DOF. (2014). (Diario Oficial de la Federación). Reglamento de la Ley General del Equilibrio Ecológico y la Protección al Ambiente en materia de Áreas Naturales Protegidas (Última Reforma DOF 21-05-2014). Gobierno de México. <https://www.gob.mx/profepa/documentos/reglamento-de-la-ley-general-del-equilibrio-ecologico-en-materia-de-areas-naturales-protegidas>. Consultado el 08 de febrero de 2024.

Escobar, F., Koleff, P., & Rös, M. (2009). Evaluación de capacidades para el conocimiento: El Sistema Nacional de Información sobre Biodiversidad como un estudio de caso. In *México: Capacidades para la conservación y el uso sustentable de la biodiversidad* (pp. 23–49). Comisión Nacional para el Conocimiento y Uso de la Biodiversidad y Programa de las Naciones Unidas para el Desarrollo.

GBIF. (2023). (*Global Biodiversity Information Facility*) *Country report—Mexico* (p. 7). Global Biodiversity Information Facility. <https://www.gbif.org/es/country/MX/summary>

Gómez-Moreno, V. del C., González-Gaona, O. J., López-Mancilla, A., Montoya-Cruz, L., Vela-Puga, J. J., Niño-Maldonado, S., Gómez-Moreno, V. del C., González-Gaona, O. J., López-Mancilla, A., Montoya-Cruz, L., Vela-Puga, J. J., & Niño-Maldonado, S. (2022). Dinámica estacional de las comunidades de aves en el matorral submontano de la Sierra de San Carlos, Tamaulipas, México. *Acta zoológica mexicana*, 38. <https://doi.org/10.21829/azm.2022.3912416>

Gorleri, F. C., Jordan, E. A., Roesler, I., Monteleone, D., & Areta, J. I. (2023). Using photographic records to quantify accuracy of bird identifications in citizen science data. *Ibis*, 165(2), 458–471. <https://doi.org/10.1111/ibi.13137>

Groom, Q., Weatherdon, L., & Geijzendorffer, I. R. (2017). Is citizen science an open science in the case of biodiversity observations? *Journal of Applied Ecology*, 54(2), 612–617. <https://doi.org/10.1111/1365-2664.12767>

Hughes, A. C., Orr, M. C., Ma, K., Costello, M. J., Waller, J., Provoost, P., Yang, Q., Zhu, C., & Qiao, H. (2021). Sampling biases shape our view of the natural world. *Ecography*, 44(9), 1259–1269. <https://doi.org/10.1111/ecog.05926>

Jiménez, R., Koleff, P., Urquiza-Haas, T., Ruíz González, S. P., Sánchez, C., Oliveros, O., Burgeff, C., Acevedo Gasman, F., Lara Morales, L., Cerdeira-Estrada, S., Careaga Olvera, S., Lozano Ramos, T., Sánchez Avelar, M., Rodríguez, P., Fernández, C., Martínez, J. M., Marín, A., Moreno, E., Villalón, R., ... Mohar, M. (2016). La informática de la biodiversidad: Una herramienta para la toma de decisiones. In J. Sarukhán & I. Pisanty (Eds.), *Capital natural de México Vol. IV: Capacidades humanas e institucionales* (Vol. 4, pp. 143–191). CONABIO.

Jin, J., & Yang, J. (2020). BDcleaner: A workflow for cleaning taxonomic and geographic errors in occurrence data archived in biodiversity databases. *Global Ecology and Conservation*, 21, e00852. <https://doi.org/10.1016/j.gecco.2019.e00852>

Kamp, J., Oppel, S., Heldbjerg, H., Nyegaard, T., & Donald, P. F. (2016). Unstructured citizen science data fail to detect long-term population declines of common birds in Denmark. *Diversity and Distributions*, 22(10), 1024–1035. <https://doi.org/10.1111/ddi.12463>

MacGregor-Fors, I., Escobar, F., Escobar-Ibáñez, J. F., Mesa-Sierra, N., Alvarado, F., Rueda-Hernández, R., Moreno, C. E., Falfán, I., Corro, E. J., Pineda, E., Bourg, A., Aguilar-López, J. L., &

- Dáttilo, W. (2022). Shopping for Ecological Indices? On the Use of Incidence-Based Species Compositional Similarity Measures. *Diversity*, 14(5), Article 5. <https://doi.org/10.3390/d14050384>
- MacGregor-Fors, I., Gómez-Martínez, M. A., Vázquez, L.-B., & Martínez, M. L. (2022). Birds of the Land of Swallows: Contribution of the main ecosystems of Cozumel Island to its avian diversity. *Écoscience*, 29(1), 15–24. <https://doi.org/10.1080/11956860.2021.1932293>
- MacGregor-Fors, I., & Schondube, J. E. (2011). Use of Tropical Dry Forests and Agricultural Areas by Neotropical Bird Communities. *Biotropica*, 43(3), 365–370. <https://doi.org/10.1111/j.1744-7429.2010.00709.x>
- Magurran, A. E. (2004). *Measuring biological diversity*. Wiley Blackwell.
- Maldonado, C., Molina, C. I., Zizka, A., Persson, C., Taylor, C. M., Albán, J., Chilquillo, E., Rønsted, N., & Antonelli, A. (2015). Estimating species diversity and distribution in the era of Big Data: To what extent can we trust public databases? *Global Ecology and Biogeography*, 24(8), 973–984. <https://doi.org/10.1111/geb.12326>
- Manzano-Fischer, P., List, R., Ceballos, G., & Cartron, J.-L. E. (2006). Avian diversity in a priority area for conservation in North America: The Janos-Casas Grandes Prairie Dog Complex and adjacent habitats in northwestern Mexico. *Biodiversity & Conservation*, 15(12), 3801–3825. <https://doi.org/10.1007/s10531-005-5408-7>
- Martínez-Morales, M. Á., Mendiola Islas, V., Zuria, I., Chávez Peón Hoffmann-Pinther, M. C., & Campuzano Velasco, R. G. (2013). La conservación de las aves más allá de las áreas naturales protegidas: El caso de la avifauna del Rancho Santa Elena, Hidalgo. *Huitzil*, 14(2), 87–100.
- Mellink, E. (1991). Bird communities associated with three traditional agroecosystems in the San Luis Potosi Plateau, Mexico. *Agriculture, Ecosystems & Environment*, 36(1), 37–50. [https://doi.org/10.1016/0167-8809\(91\)90034-U](https://doi.org/10.1016/0167-8809(91)90034-U)
- Navarro-Sigüenza, A. G., Rebón-Gallardo, Ma. F., Gordillo-Martínez, A., Townsend Peterson, A., Berlanga-García, H., & Sánchez-González, L. A. (2014). Biodiversidad de aves en México. *Revista Mexicana de Biodiversidad*, 85, 476–495.
- R Core Team. (2023). *R: A language and environment for statistical computing* (R version 4.3.0 (2023-04-21) -- “Already Tomorrow”) [X86_64-pc-linux-gnu (64-bit)]. R Foundation for Statistical Computing. <http://www.R-project.org>
- Ralph, C. J., Geupel, G. R., Pyle, P., Martin, T. E., DeSante, D. F., & Milá, B. (1996). Manual de métodos de campo para el monitoreo de aves terrestres. (General Technical Report PSW-GTR-159-Web; p. 46). Southwest Research Station, Forest Service, U.S. Department of Agriculture.
- Ramírez-Albores, J. E. (2006). Variación en la composición de comunidades de aves en la Reserva de la Biosfera Montes Azules y áreas adyacentes, Chiapas, México. *Biota Neotropica*, 6. <https://doi.org/10.1590/S1676-06032006000200019>
- Ramírez-Albores, J. E. (2007). Bird diversity and conservation of Alto Balsas (Southwestern Puebla), Mexico. *Revista de Biología Tropical*, 55(1), 287–300. <https://doi.org/10.15517/rbt.v55i1.6081>

- Ramírez-Albores, J. E. (2010). Diversidad de aves de hábitats naturales y modificados en un paisaje de la Depresión Central de Chiapas, México. *Revista de Biología Tropical*, 58(1), 511–528. <https://doi.org/10.15517/rbt.v58i1.5225>
- Ramírez-Albores, J. E. (2013). Riqueza y diversidad de aves de un área de la Faja Volcánica Transmexicana, Tlaxcala, México. *Acta Zoológica Mexicana*, 29(3), 486–512. <https://doi.org/10.21829/azm.2013.2931593>
- Ramírez-Albores, J. E., Pérez-Suárez, M., Ramírez-Albores, J. E., & Pérez-Suárez, M. (2018). Tropical forest remnants as shelters of avian diversity within a tourism development matrix in Yucatan Peninsula, Mexico. *Revista de Biología Tropical*, 66(2), 799–813. <https://doi.org/10.15517/rbt.v66i2.33410>
- Randler, C. (2021). Users of a citizen science platform for bird data collection differ from other birdwatchers in knowledge and degree of specialization. *Global Ecology and Conservation*, 27, e01580. <https://doi.org/10.1016/j.gecco.2021.e01580>
- Ribeiro, B. R., Velazco, S., Guidoni-Martins, K., Tessarolo, G., Jardim, L., Bachman, S., & Loyola, R. (2022a). *bdc: Biodiversity Data Cleaning* (1.1.2) [Computer software]. <https://CRAN.R-project.org/package=bdc>
- Ribeiro, B. R., Velazco, S. J. E., Guidoni-Martins, K., Tessarolo, G., Jardim, L., Bachman, S. P., & Loyola, R. (2022b). *bdc: A toolkit for standardizing, integrating and cleaning biodiversity data*. *Methods in Ecology and Evolution*, 13(7), 1421–1428. <https://doi.org/10.1111/2041-210X.13868>
- Rojas-Soto, O. R., Oliveras de Ita, A., Almazán-Núñez, R. C., Navarro-Sigüenza, A. G., & Sánchez-González, L. A. (2009). Avifauna de Campo Morado, Guerrero, México. *Revista Mexicana de Biodiversidad*, 80(3), 741–749. <https://doi.org/10.22201/ib.20078706e.2009.003.170>
- Rosenblatt, C. J., Dayer, A. A., Duberstein, J. N., Phillips, T. B., Harshaw, H. W., Fulton, D. C., Cole, N. W., Raedeke, A. H., Rutter, J. D., & Wood, C. L. (2022). Highly specialized recreationists contribute the most to the citizen science project eBird. *Ornithological Applications*, 124(2), duac008. <https://doi.org/10.1093/ornithapp/duac008>
- Ruiz-Campos, G., & Rodríguez-Meraz, M. (1997). Composición taxonómica y ecológica de la avifauna de los ríos El Mayor y Hardy, y áreas adyacentes en el Valle de Mexicali, Baja California, México. *Anales Del Instituto de Biología. Serie Zoología*, 68(2), 291–315.
- Sahagún-Sánchez, F. J., Espinoza-Guzmán, M. A., & Huerta-Martínez, F. M. (2021). Riqueza de aves en la Reserva de la Biosfera Sierra del Abra Tanchipa en San Luis Potosí, México. *CienciaUAT*, 15(2), 37–67. <https://doi.org/10.29059/cienciauat.v15i2.1439>
- Santiago-Alarcon, D., Arriaga-Weiss, S. L., & Escobar, O. (2011). Bird community composition of Centla Marshes Biosphere Reserve, Tabasco, Mexico. *Ornitología Neotropical*, 22, 229–246.
- SEDEMA. (2023a). (Secretaría de Medio Ambiente del Estado de Veracruz). Estudio Previo Justificativo “Reserva Ecológica La Cortadura” (p. 442) [Estudio Técnico Justificativo]. Secretaría de Medio Ambiente del Estado de Veracruz, H. Ayuntamiento de Coatepec, Veracruz 2022-2025, Pronatura Veracruz A.C.

- SEDEMA. (2023b). (Secretaría de Medio Ambiente del Estado de Veracruz). Instructivo para elaborar la “Manifestación de Impacto Ambiental para proyectos de fraccionamientos, lotificaciones y unidades habitacionales.” Gobierno del Estado de Veracruz. <http://www.veracruz.gob.mx/medioambiente/formatos-impacto-ambiental/>. Consultado el 08 de febrero de 2024.
- SEMARNAT. (2002). (Secretaria de Medio Ambiente y Recursos Naturales). Guía para la presentación de la Manifestación de Ambiental (sic.) del sector eléctrico. Modalidad: Particular (1st ed.). Secretaria de Medio Ambiente y Recursos Naturales. <https://www.gob.mx/semarnat/documentos/guias-de-impacto-ambiental>. Consultado el 08 de febrero de 2024.
- SEMARNAT. (2017). (Secretaria de Medio Ambiente y Recursos Naturales). Guía para la elaboración de Estudios Técnicos Justificativos. Cambio de Uso de Suelo en Terrenos Forestales (CUSTF). Secretaria de Medio Ambiente y Recursos Naturales. https://www.gob.mx/cms/uploads/attachment/file/233657/Guia_Final_Estudio_Tecnico_Justificativo.pdf. Consultado el 08 de febrero de 2024.
- Soares, L., Cockle, K. L., Ruelas Inzunza, E., Ibarra, J. T., Miño, C. I., Zuluaga, S., Bonaccorso, E., Ríos-Orjuela, J. C., Montañó-Centellas, F. A., Freile, J. F., Echeverry-Galvis, M. A., Bonaparte, E. B., Diele-Viegas, L. M., Speziale, K., Cabrera-Cruz, S. A., Acevedo-Charry, O., Velarde, E., Cuatianquiz Lima, C., Ojeda, V. S., ... Martins, P. V. R. (2023). Neotropical ornithology: Reckoning with historical assumptions, removing systemic barriers, and reimagining the future. *Ornithological Applications*, duac046. <https://doi.org/10.1093/ornithapp/duac046>
- Soberón, J. (2022). Biodiversity Informatics for Public Policy. The case of CONABIO in Mexico. *Biodiversity Informatics*, 17. <https://journals.ku.edu/jbi/article/view/18270>
- Soberón, J., & Peterson, T. (2004). Biodiversity informatics: Managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359(1444), 689–698. <https://doi.org/10.1098/rstb.2003.1439>
- Swanson, A., Kosmala, M., Lintott, C., & Packer, C. (2016). A generalized approach for producing, quantifying, and validating citizen science data from wildlife images. *Conservation Biology*, 30(3), 520–531. <https://doi.org/10.1111/cobi.12695>
- Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R., & Legendre, F. (2017). Taxonomic bias in biodiversity data and societal preferences. *Scientific Reports*, 7(1), 9132. <https://doi.org/10.1038/s41598-017-09084-6>
- Ugalde-Lezama, S., Alcántara-Carbajal, J. L., Valdez-Hernández, J. I., Ramírez-Valverde, G., Velázquez-Mendoza, J., & Tarángo-Arámbula, L. A. (2010). Riqueza, abundancia y diversidad de aves en un bosque templado con diferentes condiciones de perturbación. *Agrociencia*, 44(2), 159–169.
- Urquiza-Haas, E. (2016). La ciencia ciudadana, una estrategia central para fortalecer e impulsar el conocimiento de la biodiversidad del país (recuadro 9.3 en Capítulo 9). In *Capital Natural de México Vol. IV: Capacidades humanas, institucionales y financieras*. (pp. 319–323). Comisión Nacional para el Conocimiento y Uso de la Biodiversidad. México.
- Urquiza-Haas, E., Koleff, P., & Ruiz González, S. P. (2016). Algunas organizaciones, redes e iniciativas relacionadas con la informática de la biodiversidad en el mundo (recuadro 5.1 en Capítulo 5). In

Capital Natural de México Vol. IV: Capacidades humanas, institucionales y financieras. (pp. 145–151). Comisión Nacional para el Conocimiento y Uso de la Biodiversidad. México.

Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T., & Vieglais, D. (2012). Darwin Core: An evolving community-developed biodiversity data standard. *PLOS ONE*, 7(1), e29715. <https://doi.org/10.1371/journal.pone.0029715>

Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Edler, D., Farooq, H., Herdean, A., Ariza, M., Scharn, R., Svantesson, S., Wengström, N., Zizka, V., & Antonelli, A. (2019). CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution*, 10(5), 744–751. <https://doi.org/10.1111/2041-210X.13152>

Zizka, A., Carvalho, F. A., Calvente, A., Baez-Lizarazo, M. R., Cabral, A., Coelho, J. F. R., Colli-Silva, M., Fantinati, M. R., Fernandes, M. F., Ferreira-Araújo, T., Moreira, F. G. L., Santos, N. M. C., Santos, T. A. B., Santos-Costa, R. C. dos, Serrano, F. C., Silva, A. P. A. da, Soares, A. de S., Souza, P. G. C. de, Tomaz, E. C., ... Antonelli, A. (2020). No one-size-fits-all solution to clean GBIF. *PeerJ*, 8, e9916. <https://doi.org/10.7717/peerj.9916>