

1 **Chromosomal-level genome assembly of golden birdwing *Troides aeacus* (Felder &  
2 Felder, 1860)**

3 Hong Kong Biodiversity Genomics Consortium

4

5 Project Coordinator and Co-Principal Investigators: Jerome H.L. Hui<sup>1</sup>, Ting Fung Chan<sup>2</sup>,  
6 Leo L. Chan<sup>3</sup>, Siu Gin Cheung<sup>4</sup>, Chi Chiu Cheang<sup>5,6</sup>, James K.H. Fang<sup>7</sup>, Juan D. Gaitan-  
7 Espitia<sup>8</sup>, Stanley C.K. Lau<sup>9</sup>, Yik Hei Sung<sup>10,11</sup>, Chris K.C. Wong<sup>12</sup>, Kevin Y.L. Yip<sup>13,14</sup>,  
8 Yingying Wei<sup>15</sup>

9 DNA & RNA extraction, library preparation and sequencing: Wai Lok So<sup>1</sup>

10 Genome assembly and gene model prediction: Wenyen Nong<sup>1</sup>

11 Samples collector and logistics: Hydrogen S.F. Pun<sup>16</sup>, Wing Kwong Yau<sup>16</sup>, Colleen Y.L.  
12 Chiu<sup>16</sup>, Sammi S.S. Chan<sup>16</sup>, Kacy K.L. Man<sup>16</sup>, Ho Yin Yip<sup>1</sup>

13

14 1. School of Life Sciences, Simon F.S. Li Marine Science Laboratory, State Key Laboratory  
15 of Agrobiotechnology, Institute of Environment, Energy and Sustainability, The Chinese  
16 University of Hong Kong, Hong Kong, China

17 2. School of Life Sciences, State Key Laboratory of Agrobiotechnology, The Chinese  
18 University of Hong Kong, Hong Kong SAR, China

19 3. State Key Laboratory of Marine Pollution and Department of Biomedical Sciences, City  
20 University of Hong Kong, Hong Kong SAR, China

21 4. State Key Laboratory of Marine Pollution and Department of Chemistry, City University of  
22 Hong Kong, Hong Kong SAR, China

23 5. Department of Science and Environmental Studies, The Education University of Hong  
24 Kong, Hong Kong SAR, China

25 6. EcoEdu PEI, Charlottetown, PE, C1A 4B7, Canada

26 7. Department of Food Science and Nutrition, Research Institute for Future Food, and State  
27 Key Laboratory of Marine Pollution, The Hong Kong Polytechnic University, Hong Kong  
28 SAR, China

29 8. The Swire Institute of Marine Science and School of Biological Sciences, The University  
30 of Hong Kong, Hong Kong SAR, China

31 9. Department of Ocean Science, The Hong Kong University of Science and Technology,  
32 Hong Kong SAR, China

33 10. Science Unit, Lingnan University, Hong Kong SAR, China

34 11. School of Allied Health Sciences, University of Suffolk, Ipswich, IP4 1QJ, UK

35 12. Croucher Institute for Environmental Sciences, and Department of Biology, Hong Kong  
36 Baptist University, Hong Kong SAR, China

37 13. Department of Computer Science and Engineering, The Chinese University of Hong  
38 Kong, Hong Kong SAR, China

39 14. Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA, USA.

40 15. Department of Statistics, The Chinese University of Hong Kong, Hong Kong SAR, China

41 16. Fung Yuen Butterfly Reserve, Hong Kong SAR, China

42

43 Correspondence on behalf of the consortium: jeromehui@cuhk.edu.hk

44

45

46 **Abstract**

47 *Troides aeacus*, the golden birdwing (Lepidoptera, Papilionidae) is a large swallowtail  
48 butterfly widely distributed in Asia. Despite its occurrence, *T. aeacus* has been assigned as a  
49 major protective species in many places given the loss of their native habitats under  
50 urbanisation and anthropogenic activities. Nevertheless, the lack of its genomic resources  
51 hinders our understanding of their biology, diversity, as well as carrying out conservation  
52 measures based on genetic information or markers. Here, we report the first chromosomal-  
53 level genome assembly of *T. aeacus* using a combination of PacBio SMRT and Omni-C  
54 scaffolding technologies. The assembled genome (351 Mb) contains 98.94% of the sequences  
55 anchored to 30 pseudo-molecules. The genome assembly also has high sequence continuity  
56 with scaffold length N50 = 12.2 Mb. A total of 28,749 protein-coding genes were predicted,  
57 and high BUSCO score completeness (98.9% of BUSCO metazoa\_odb10 genes) was also  
58 revealed. This high-quality genome offers a new and significant resource for understanding  
59 the swallowtail butterfly biology, as well as carrying out conservation measures of this  
60 ecologically important lepidopteran species.

61

62 **Introduction**

63 The golden birdwing butterfly *Troides aeacus* (Figure 1A) is a swallowtail butterfly  
64 that are widely distributed in Asia, including Bangladesh, Burma, Cambodia, China, India,  
65 Laos, Malaysia, Nepal, Thailand, and Vietnam (Böhm et al 2020). The species is generally  
66 large in size with wingspan reaching ~15 cm, and has iconic black forewings and golden-  
67 yellow hindwings carved with grey stripes and black spots (Wu et al 2010; Li et al 2010).  
68 Due to its attractive appearances, it has also been vastly collected and traded in curio markets  
69 (Collins and Morris 1985, New and Collins 1991; Wu et al 2010).

70 Similar to other homometabolans, *T. aeacus* exhibits larvae and pupae stages (5 larval  
71 instar stages before transforming into green girdled pupal stage)(Li et al 2010). The larvae are  
72 generally dependent on host plants in the Aristolochiaceae, especially of the genus  
73 *Aristolochia* that can be commonly found in Asia (Li et al 2010; Wu et al 2010; Böhm et al  
74 2020; Chen et al 2013). After emergence, the adults will feed and live around nectaring  
75 flowers such as those in the genus *Hibiscus*, *Ixora*, *Lantana*, *Mussaenda*, and *Spathodea*  
76 (Cotton and Racheli 2006; Böhm et al 2020). Anthropogenic activities including deforestation,  
77 grazing, herbicide application, hunting, land reclamation, mine exploitation, and trading all  
78 have been suggested to pose threats to *T. aeacus* (Böhm et al 2020; Li et al 2010; Khanal  
79 2022). In certain places such as Hong Kong, *T. aeacus* has also been suggested to be under  
80 protection and restoration of its lost; and in Taiwan, the trade of endemic subspecies such as *T.*  
81 *aeacus* is protected by the Convention on International Trade in Endangered Species of Wild  
82 Fauna and Flora (CITES) (Fellowes et al 2002).

83

84 **Context**

85 To date, genomic resource in the genus *Troides* are confined to *T. helena* (He et al  
86 2022) and *T. oblongomaculatus* (Reboud et al 2023). In light of the high conservation value  
87 of *T. aeacus* and its phylogenetic position to understand the diversification of butterflies  
88 (Condamine et al 2015), it has been chosen as one of the species for genome sequencing by  
89 the Hong Kong Biodiversity Genomics Consortium (a.k.a. EarthBioGenome Project Hong

90 Kong), which is formed by investigators from 8 publicly funded universities in Hong Kong.  
91 Here, we report a chromosomal-level genome assembly of the golden birdwing *T. aeacus*.

92

### 93 **Methods**

#### 94 *Sample collection and species identification*

95 A pupa of the golden birdwing *Troides aeacus* was obtained in August 2022. The pupa  
96 was snap-frozen in liquid nitrogen upon collection. The frozen pupa was further grinded into  
97 fine powder and stored at -80 °C freezer until DNA isolation. A portion of the powder was  
98 used for species molecular identification with QIAamp DNA Mini Kit (Qiagen Cat. 51306),  
99 following the provided protocol. The DNA was then used as a template for conventional PCR  
100 with the following protocol: an initial denaturation step at 95 °C for 3 minutes; followed by  
101 36 amplification cycles of 30 seconds for denaturation at 95 °C; 30 seconds for primer  
102 annealing at 55 °C and 1 minute for extension at 72 °C; and a final extension step at 72 °C for  
103 3 minutes. The reaction mixture included PCR buffer, DNA template, 2 mM dNTP, 1.5 mM  
104 MgCl<sub>2</sub>, 0.4 mM of each forward and reverse primers (LCO1490: 5'-  
105 GGTCAACAAATCATAAAGATATTGG-3', HCO2198: 5'-  
106 TAAACTTCAGGGTGACCAAAAAATCA-3') (Folmer et al 1994), and *Taq* DNA  
107 polymerase. The PCR was performed on a T100™ thermal cycler (Bio-Rad, USA).

108

#### 109 *Isolation of high molecular weight genomic DNA*

110 High molecular weight genomic DNA was isolated with the remaining stored powder  
111 using the Qiagen MagAttract HMW kit (Qiagen Cat. No. 67563) following the protocol. In  
112 summary, 1 g of sample powder was placed in microcentrifuge tube with 200 µl 1X  
113 phosphate-buffered saline (PBS), RNase A, Proteinase K, and Buffer AL. The mixture was  
114 incubated at room temperature (22-25 °C) for 2.5 hours, until the tissue was completely  
115 disintegrated. The sample was eventually eluted with 120 µl of elution buffer (PacBio Ref.  
116 No. 101-633-500) and stored at 4 °C. In order to keep the integrity of the DNA, wide-bore  
117 pipette tips were used in any transfer of DNA during the process. The sample was then  
118 proceeded to quality control by quantifying with the Qubit® Fluorometer and Qubit™ dsDNA  
119 HS and BR Assay Kits (Invitrogen™ Cat. No. Q32851). An overnight pulse-field gel  
120 electrophoresis was performed to estimate the size of the isolated DNA, with three markers  
121 (λ-Hind III digest; Takara Cat. No. 3403, DL15,000 DNA Marker; Takara Cat. No. 3582A  
122 and CHEF DNA Size Standard-8-48 kb Ladder; Cat. No. 170-3707). Besides, the sample  
123 purity was examined by the NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer  
124 (with A260/A280: ~1.8 and A260/A230: >2.0 as a standard threshold).

125

#### 126 *DNA shearing, PacBio library preparation and sequencing*

127 120 µl of DNA sample of 10 µg DNA was transferred to a g-tube (Covaris Part No.  
128 520079). The tube was then allowed to proceed 6 passes of centrifugation with 2,000 x g of 2  
129 minutes each. The resultant DNA was saved in a 2 mL DNA LoBind® Tube (Eppendorf Cat.  
130 No. 022431048) at 4 °C until library preparation. The molecular weight of the isolated DNA  
131 was examined by overnight pulse-field gel electrophoresis. The electrophoresis profile was  
132 set as follow: 5K as the lower end and 100K as the higher end for the designated molecular  
133 weight; Gradient = 6.0V/cm; Run time = 15 h:16 min; included angle = 120 °; Int. Sw. Tm =

134 22 s; Fin. Sw. Tm = 0.53 s; Ramping factor: a = Linear. The gel was run in 1.0% PFC agarose  
135 in 0.5X TBE buffer at 14 °C.

136 A SMRTbell library was made using the SMRTbell® prep kit 3.0 (PacBio Ref. No.  
137 102-141-700), following the provided protocol. In summary, single-stranded overhangs of the  
138 genomic DNA were removed, and the DNA is repaired from physical damage during shearing.  
139 Subsequently, both DNA ends were tailed with an A-overhang and ligation of T-overhang  
140 SMRTbell adapters was performed at 20 °C for 30 minutes. The SMRTbell library was then  
141 purified with SMRTbell® cleanup beads (PacBio Ref. No. 102158-300). The size and  
142 concentration of the library were assessed with the pulse-field gel electrophoresis and the  
143 Qubit® Fluorometer and Qubit™ dsDNA HS and BR Assay Kits (Invitrogen™ Cat. No.  
144 Q32851), respectively. A subsequent nuclease treatment step was carried out to remove non-  
145 SMRTbell structures in the library. A final size-selection step was performed to remove small  
146 DNA fragments in the library with 35% AMPure PB beads. The Sequel® II binding kit 3.2  
147 (PacBio Ref. No. 102-194-100) was used for final preparation. In short, Sequel II primer 3.2  
148 and Sequel II DNA polymerase 2.2 were annealed and bound to the SMRTbell library,  
149 respectively. The library was loaded afterwards at an on-plate concentration of 50-90 pM  
150 using the diffusion loading mode. The sequencing was conducted on the Sequel IIe System  
151 with an internal control provided in the kit. The sequencing was performed in 30-hour movies,  
152 with 120 min pre-extension, connected to the software SMRT Link v11.0 (PacBio). HiFi  
153 reads are eventually generated and collected for further analysis. One SMRT cell was used for  
154 this sequencing (Supplementary Information 1).

155

### 156 *Omni-C library preparation and sequencing*

157 An Omni-C library was made using the Dovetail® Omni-C® Library Preparation  
158 Kit (Dovetail Cat. No. 21005) according to the provided protocol. In summary, 80 mg of  
159 frozen powdered tissue sample was placed in a microcentrifuge tube with 1 mL 1X PBS and  
160 formaldehyde. The fixed DNA was digested with endonuclease DNase I. Afterwards, the  
161 concentration and size of the digested sample was examined by the Qubit® Fluorometer and  
162 Qubit™ dsDNA HS and BR Assay Kits (Invitrogen™ Cat. No. Q32851), and the TapeStation  
163 D5000 HS ScreenTape, respectively. Both DNA ends were the polished and ligation of  
164 biotinylated bridge adaptor was conducted at 22 °C for 30 minutes. Subsequent proximity  
165 ligation between crosslinked DNA was performed at 22 °C for 1 hour. After ligation, the  
166 DNA was reverse crosslinked, and purified with SPRIselect™ Beads (Beckman Coulter  
167 Product No. B23317) to remove the biotin that was not internal to the ligated fragments. The  
168 Dovetail™ Library Module for Illumina (Dovetail Cat. No. 21004) was used for end repair  
169 and adapter ligation. The DNA was tailed with an A-overhang, which allowed Illumina-  
170 compatible adapters to ligate to the DNA fragments at 20 °C for 15 minutes. The Omni-C  
171 library was then sheared into fragments with USER Enzyme Mix and further purified with  
172 SPRIselect™ Beads. Isolation of DNA fragments with internal biotin were performed with  
173 Streptavidin Beads. Universal and Index PCR Primers from the Dovetail™ Primer Set for  
174 Illumina (Dovetail Cat. No. 25005) were used to amplify the constructed library. Size  
175 selection was carried out with SPRIselect™ Beads targeting fragments ranging between 350  
176 bp and 1000 bp. Finally, the concentration and fragment size of the sequencing library was  
177 examined with the Qubit® Fluorometer and Qubit™ dsDNA HS and BR Assay Kits, and the  
178 TapeStation D5000 HS ScreenTape, respectively. The resultant library was sequenced on the  
179 Illumina HiSeq-PE150 platform (Supplementary Information 1).

180

181 *Genome assembly and gene model prediction*

182 *De novo* genome assembly was performed using Hifiasm (Cheng et al 2021).  
183 Haplotypic duplications were identified and removed using purge\_dups based on the depth of  
184 HiFi reads (Guan et al 2020). Proximity ligation data from the Omni-C library were used to  
185 scaffold genome assembly by YaHS (Zhou et al., 2022). Transposable elements (TEs) were  
186 annotated as previously described (Baril et al 2022), using the automated Earl Grey TE  
187 annotation pipeline (version 1.2, <https://github.com/TobyBaril/EarlGrey>). Gene models were  
188 predicted from the soft-masked genome assembled by funannotate (Palmer & Stajich, 2020)  
189 using predict with the parameters "--protein\_evidence uniprot\_sprot.fasta --genemark\_mode  
190 ES --optimize\_augustus --busco\_db metazoa --organism other --max\_intronlen 350000" and  
191 other default parameters.

192

193 **Results and discussion**

194 *Genome assembly of *T. aeacus**

195 A total of 27 Gb of HiFi bases were yielded with an average HiFi read length of 9,688  
196 bp with 78X coverage (Supplementary Information 1). After incorporating with 21.7 Gb  
197 Omni-C data, the resulting genome assembly was 350.66 Mb in size with 36 scaffolds, of  
198 which 30 scaffolds are of chromosome length (Figure 1B-C; Table 1-2). The genome has a  
199 high contiguity with a scaffold N50 value of 12.21 Mb and high completeness with the  
200 complete BUSCO estimation to be 98.9% (metazoa\_odb10)(Figure 1B, Table 1). While the  
201 genome size estimation was about 268.3 Mb with 2.93% nucleotide heterozygosity rate  
202 (Figure 1D; Supplementary Information 2), the assembled *Troides aeacus* genome has a  
203 genome size similar to other swallowtail butterfly genomes, including *Troides Helena* (~330  
204 Mb) (He et al 2022) and *Troides oblongomaculatus* (~348 Mb) (Reboud et al 2023). In  
205 addition, 43 telomeres were found in 25 scaffolds of the assembly genome (Table 3).  
206 Furthermore, 32,183 gene models were predicted with a BUSCO score of 86.5%.

207

208 *Repeat content*

209 A total repetitive content of 29.50% was identified in the assembled genome,  
210 including 5.16% unclassified elements (Figure 1E; Table 4). Among the known repeats, LINE  
211 is the most abundant (12.01%), followed by SINE retrotransposons (6.38%) and DNA  
212 transposons (4.71%), whereas Rolling Circle, LTR, Penelope and other are present in low  
213 proportions (Rolling Circle: 0.78%, LTR: 0.26%, Penelope: 0.20%, other: 0.02%).

214

215 **Conclusion and reuse potential**

216 This study presents the first chromosomal-level genome assembly of the golden  
217 birdwing *Troides aeacus*, which is a useful and precious resource for further phylogenomic  
218 studies of birdwing butterfly species in light of species diversification and conservation.

219

220 **Data validation and quality control**

221 During DNA extraction and PacBio library preparation, the samples were subjected  
222 to quality control with NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer,

223 Qubit® Fluorometer, and overnight pulse-field gel electrophoresis. The Omni-C library was  
224 inspected by Qubit® Fluorometer and TapeStation D5000 HS ScreenTape.

225 Regarding the genome assembly, the Hifiasm output was blast to the NT database and  
226 the resultant output was used as the input for BlobTools (v1.1.1) (Laetsch & Blaxter 2017).  
227 Scaffolds that were identified as possible contamination were removed from the assembly  
228 manually (Supplementary Information 3). A statistical kmer-based approach was applied to  
229 estimate the heterozygosity of the assembled genome heterozygosity. The repeat content and  
230 the corresponding sizes were analysed using Jellyfish (Marçais & Kingsford 2011) and  
231 GenomeScope (Ranallo-Benavidez et al 2020) (Figure 1D; Supplementary Information 2).  
232 Furthermore, telomeric repeats were inspected by FindTelomeres  
233 (<https://github.com/JanaSperschneider/FindTelomeres>). Benchmarking Universal Single-  
234 Copy Orthologs (BUSCO, v5.5.0) (Manni et al., 2021) was used to assess the completeness  
235 of the genome assembly and gene annotation with metazoan dataset (metazoa\_odb10).

236

## 237 **Data availability**

238 The final genome assembly was submitted to NCBI under accession numbers  
239 GCA\_033220335.2 ([https://www.ncbi.nlm.nih.gov/datasets/genome/GCA\\_033220335.2](https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_033220335.2)).  
240 The raw reads yielded from this study were deposited to the NCBI database under the  
241 BioProject accession number PRJNA973839. The genome annotation files were deposited in  
242 the figshare (<https://figshare.com/s/811f9d4158e592e9d209>).

243

## 244 **Funding**

245 This work was funded and supported by the Hong Kong Research Grant Council  
246 Collaborative Research Fund (C4015-20EF), CUHK Strategic Seed Funding for  
247 Collaborative Research Scheme (3133356) and CUHK Group Research Scheme (3110154).

248

## 249 **Author's contributions**

250 JHLH, TFC, LLC, SGC, CCC, JKHF, JDG, SCKL, YHS, CKCW, KYLY and YW conceived  
251 and supervised the study; WLS carried out DNA extraction, library preparation and  
252 sequencing; WN performed genome assembly and gene model prediction; HSFP, WKY,  
253 CYLC, SSSC and KKLM and HYY collected and maintained the butterfly samples. All  
254 authors approved the final version of the manuscript.

255

## 256 **Competing interest**

257 The authors declare that they do not have competing interests.

258

## 259 **Acknowledgements**

260 The authors thank Sean Law for discussion.

261

## 262 **References**

263

264 1. Baril T, Imrie RM, Hayward A. Earl Grey: a fully automated user-friendly  
265 transposable element annotation and analysis pipeline. bioRxiv. 2022.  
266 <https://doi.org/10.1101/2022.06.30.498289>

267 2. Böhm M, Chowdhury S, Khanal B, Lo P, Monastyrskii A. *Troides aeacus* (amended  
268 version of 2018 assessment). The IUCN Red List of Threatened Species. 2020.

269 3. Chen PJ, Arikawa K, Yang EC. Diversity of the photoreceptors and spectral  
270 opponency in the compound eye of the Golden Birdwing, *Troides aeacus formosanus*.  
271 PLoS One. 2013;8(4):e62240.

272 4. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo  
273 assembly using phased assembly graphs with hifiasm. Nature methods.  
274 2021;18(2):170-5.

275 5. Collins NM, Morris MG. Threatened swallowtail butterflies of the world: the IUCN  
276 Red Data Book. Iucn; 1985.

277 6. Condamine FL, Toussaint EF, Clamens AL, Genson G, Sperling FA, Kergoat GJ.  
278 Deciphering the evolution of birdwing butterflies 150 years after Alfred Russel  
279 Wallace. Scientific Reports. 2015;5(1):11860.

280 7. Cotton AM, Racheli T. A preliminary annotated checklist of the Papilionidae of Laos  
281 with notes on taxonomy, phenology, distribution and variation (Lepidoptera,  
282 Papilioidea). Fragmenta entomologica. 2006;38(2):279-378.

283 8. Fellowes JR, Lau MW, Dudgeon D, Reels GT, Ades GW, Carey GJ, Chan BP,  
284 Kendrick RC, Lee KS, Leven MR, Wilson KD. Wild animals to watch: Terrestrial and  
285 freshwater fauna of conservation concern in Hong Kong. Memoirs of the Hong Kong  
286 Natural History Society. 2002;25:123-59.

287 9. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. DNA primers for amplification  
288 of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates.  
289 Molecular Marine Biology and Biotechnology. 1994;3, 294–299.

290 10. Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. Identifying and  
291 removing haplotypic duplication in primary genome assemblies. Bioinformatics.  
292 2020;36(9):2896-8.

293 11. He JW, Zhang R, Yang J, Chang Z, Zhu LX, Lu SH, Xie FA, Mao JL, Dong ZW, Liu  
294 GC, Hu P. High-quality reference genomes of swallowtail butterflies provide insights  
295 into their coloration evolution. Zoological Research. 2022;43(3):367.

296 12. Khanal B. Survey of population and threats of a cites listed butterfly *Troides aeacus*  
297 *aeacus* (Felder and Felder, 1860) in central Nepal. Journal of Natural History Museum.  
298 2022:77-86.

299 13. Li XS, Luo YQ, Zhang YL, Schweiger O, Settele J, Yang QS. On the conservation  
300 biology of a Chinese population of the birdwing *Troides aeacus* (Lepidoptera:  
301 Papilionidae). Journal of Insect Conservation. 2010;14:257-68.

302 14. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel  
303 and streamlined workflows along with broader and deeper phylogenetic coverage for  
304 scoring of eukaryotic, prokaryotic, and viral genomes. Molecular biology and  
305 evolution. 2021;38(10):4647-54.

306 15. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of  
307 occurrences of k-mers. Bioinformatics. 2011;27(6):764-70.

308 16. New TR, Collins NM. Swallowtail butterflies: an action plan for their conservation.  
309 IUCN; 1991.

310 17. Palmer JM, Stajich J. Funannotate v1. 8.1: Eukaryotic genome annotation. Zenodo  
311 <https://doi.org/10.5281/zenodo.2020;4054262>.

312 18. Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for

313 reference-free profiling of polyploid genomes. *Nature communications.* 2020;11(1):1432.

314

315 19. Reboud EL, Nabholz B, Chevalier E, Tilak MK, Bito D, Condamine FL. Genomics, 316 Population Divergence, and Historical Demography of the World's Largest and 317 Endangered Butterfly, The Queen Alexandra's Birdwing. *Genome Biology and* 318 *Evolution.* 2023;15(4):evad040.

319 20. Wu, I. H., Yang, P. S., Liu, C. Y., & Yeh, W. B. Genetic differentiation of *Troides* 320 *aeacus formosanus* (Lepidoptera: Papilionidae), based on cytochrome oxidase I 321 sequences and amplified fragment length polymorphism. *Ann. Entomol. Soc. Am.*, **103**, 322 1018-1024. (2010).

323 21. Zhou, C., McCarthy, S. A., & Durbin, R. YaHS: yet another Hi-C scaffolding tool. 324 *Bioinformatics*, **39**, 1- (2023).

325

326 **Figure legends**

327 **Table 1.** Details of genome assembly statistics.

328 **Table 2.** Information of 30 chromosomal-length scaffolds.

329 **Table 3.** Summary of telomeric repeats found in 25 scaffolds.

330 **Table 4.** Summary of repetitive elements in the genome.

331

332 **Figure 1. Genomic information of *Troides aeacus*.** **A)** Photo of *T. aeacus*; **B)** Statistics of 333 the assembled genome; **C)** Omni-C contact map of the assembly visualised using Juicebox 334 (v1.11.08); **D)** Genomescope report with k-mer = 21; **E)** Repetitive elements distribution in 335 the assembled genome.

336

337 **Supplementary Information 1.** Summary of genome sequencing data.

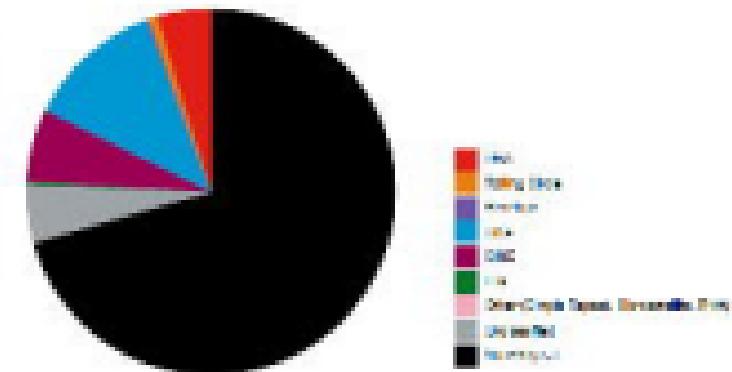
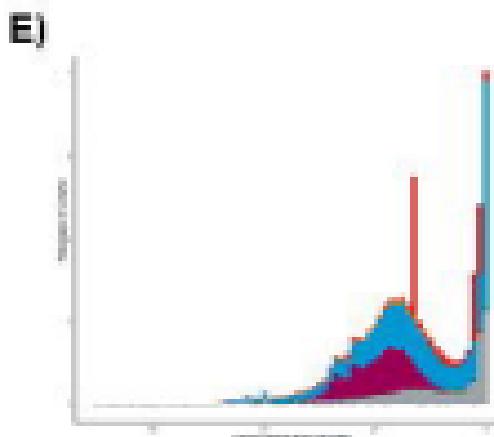
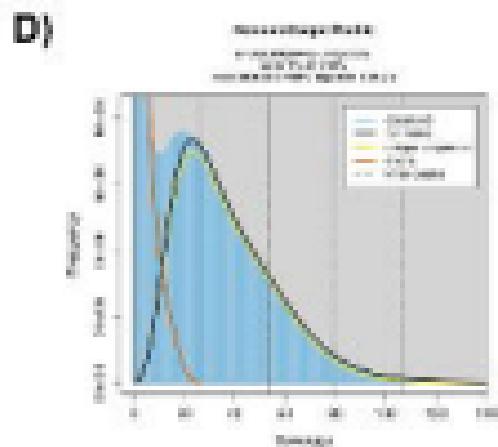
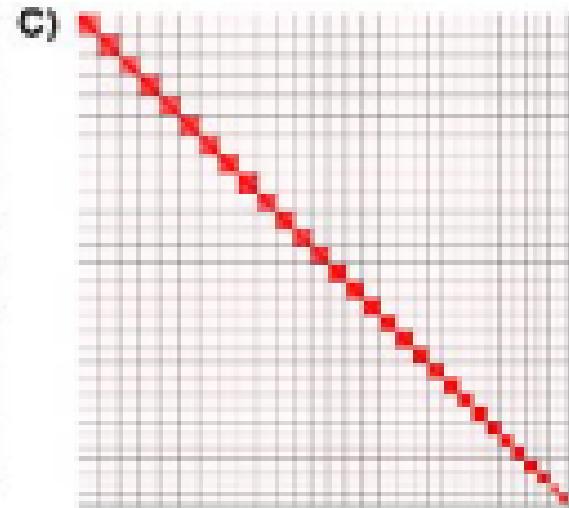
338 **Supplementary Information 2.** GenomeScope result summary (k-mer = 21).

339 **Supplementary Information 3.** Genome assembly QC and contaminant detection.



**B)**

	Trovita genome
Accession number	JV011 Statobolism
Term (BIN)	300001.300
Scaffold number	31
Mean length (bp)	9,190,915
N count	0.0000%
Gaps	4
MSB (bp)	13,212,598
BUSCOs	0.00% (0/100), 0.00% (0/10), 1.00% (1/100), 98.00% (98/100)
HIFI coverage (X)	70
Gene models	32,154



**Figure 1**