

1 BlobToolKit – Interactive quality  
2 assessment of genome assemblies

3 Richard Challis <sup>1,2\*</sup>, Edward Richards <sup>3</sup>, Jeena Rajan<sup>3</sup>, Guy Cochrane <sup>3</sup>,  
4 Mark Blaxter <sup>1,2</sup>

5  
6 1 Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, UK

7 2 Wellcome Sanger Institute, Cambridge CB10 1SA, UK

8 3 European Molecular Biology Laboratory, European Bioinformatics Institute,  
9 Wellcome Genome Campus, Cambridge CB10 1SD, UK

10

11 Supplementary File S1

12

## Figure captions and URLs

13 Most of the figures presented in this paper were generated using the BlobToolKit  
14 Viewer and can be reproduced using the following URLs. Where applicable, these  
15 are presented alongside the automatically generated captions.

### 16 Figure 1A

17 Caption: Square-binned blob plot of base coverage in SRR026696 against  
18 GC proportion for scaffolds in assembly ACVV01. Scaffolds are coloured by  
19 phylum and binned at a resolution of 30 divisions on each axis. Coloured  
20 squares within each bin are sized in proportion to the sum of individual  
21 scaffold lengths on a square-root scale, ranging from 452 to 36,531,624.  
22 Histograms show the distribution of scaffold length sum along each axis.

23 URL:

24 <https://blobtoolkit.genomehubs.org/view/ACVV01/dataset/ACVV01/blob>

### 25 Figure 1B

26 Caption: Cumulative scaffold length for assembly ACVV01. The grey line  
27 shows cumulative length for all scaffolds. Coloured lines show cumulative  
28 lengths of scaffolds assigned to each phylum using the bestsumorder  
29 taxrule.

30 URL:

31 <https://blobtoolkit.genomehubs.org/view/ACVV01/dataset/ACVV01/cumulative>

### 33 Figure 1C

34 Caption: Snail plot summary of assembly statistics for assembly ACVV01.  
35 The main plot is divided into 1,000 size-ordered bins around the  
36 circumference with each bin representing 0.1% of the 253,560,284 bp  
37 assembly. The distribution of scaffold lengths is shown in dark grey with the  
38 plot radius scaled to the longest scaffold present in the assembly  
39 (7,262,926 bp, shown in red). Orange and pale-orange arcs show the N50  
40 and N90 scaffold lengths (23,589 and 3,544 bp), respectively. The pale grey  
41 spiral shows the cumulative scaffold count on a log scale with white scale  
42 lines showing successive orders of magnitude. The blue, pale-blue and  
43 white area around the outside of the plot shows the distribution of GC, AT  
44 and N percentages in the same bins as the inner plot. A summary of  
45 complete, fragmented, duplicated and missing BUSCO genes in the  
46 diptera\_odb9 set is shown in the top right.

47 URL:  
48 <https://blobtoolkit.genomehubs.org/view/ACVV01/dataset/ACVV01/snail>

49 Figure 1D

50 URL:  
51 <https://blobtoolkit.genomehubs.org/view/ACVV01/dataset/ACVV01/busc>  
52 [o](#)

53 Figure 2:

54 A representation of the INSDC-pipeline directed acyclic graph (DAG) can be  
55 generated using Snakemake. From a local copy of the pipeline directory, a  
56 PDF format image of the DAG can be created using the command:

```
57 snakemake -p -n --configfile example.yaml --rulegraph  
58 --forceall | dot -Tpdf > DAG.pdf
```

59 Figure 3A

60 Caption: Square-binned blob plot of base coverage in SRR026696 against  
61 GC proportion for scaffolds in assembly ACVV01. Scaffolds are coloured by  
62 phylum and binned at a resolution of 30 divisions on each axis. Coloured  
63 squares within each bin are sized in proportion to the sum of individual  
64 scaffold lengths on a logarithmic scale, ranging from 867 to 40,536,114.  
65 Histograms show the distribution of scaffold length sum along each axis.

66 URL:  
67 [https://blobtoolkit.genomehubs.org/view/ACVV01/dataset/ACVV01/blob?zScale=scaleLog&catField=bestsumorder\\_phylum&otherLimit=2&palette=user&color1=rgba%28191%2C191%2C191%2C1%29&color0=rgba%28255%2C127%2C0%2C1%29&showTotal=false&bestsumorder\\_phylum--Active=true&bestsumorder\\_phylum--Order=%2CProteobacteria#Lists](https://blobtoolkit.genomehubs.org/view/ACVV01/dataset/ACVV01/blob?zScale=scaleLog&catField=bestsumorder_phylum&otherLimit=2&palette=user&color1=rgba%28191%2C191%2C191%2C1%29&color0=rgba%28255%2C127%2C0%2C1%29&showTotal=false&bestsumorder_phylum--Active=true&bestsumorder_phylum--Order=%2CProteobacteria#Lists)

72 Selection: To highlight the set of Proteobacterial scaffolds with BUSCOs,  
73 upload File S2 to the above URL then deactivate the selection using the  
74 Filters menu.

75 Figure 3B

76 Caption: Kite-shaped blob plot of base coverage in SRR026696 against GC  
77 proportion for scaffolds in assembly ACVV01. Scaffolds are coloured by  
78 phylum. Kite shapes summarise the core distribution of scaffolds.  
79 Horizontal and vertical lines represent a range spanning 2 standard  
80 deviations about the weighted mean value for each axis. The lines intersect  
81 at a point representing the weighted median value. Histograms show the  
82 distribution of scaffold length sum along each axis.

83 URL:  
84 [https://blobtoolkit.genomehubs.org/view/ACVV01/dataset/ACVV01/blob?zScale=scaleLog&catField=bestsumorder\\_phylum&otherLimit=2&palette=user&color1=rgba%28191%2C191%2C191%2C1%29&color0=rgba%28255%2C127%2C0%2C1%29&showTotal=false&bestsumorder\\_phylum--Active=true&bestsumorder\\_phylum--Order=%2CProteobacteria&plotShape=kite](https://blobtoolkit.genomehubs.org/view/ACVV01/dataset/ACVV01/blob?zScale=scaleLog&catField=bestsumorder_phylum&otherLimit=2&palette=user&color1=rgba%28191%2C191%2C191%2C1%29&color0=rgba%28255%2C127%2C0%2C1%29&showTotal=false&bestsumorder_phylum--Active=true&bestsumorder_phylum--Order=%2CProteobacteria&plotShape=kite)

### 89 Figure 3C

90 Caption: Square-binned blob plot of base coverage in SRR026696 against  
91 GC proportion for scaffolds in assembly ACVV01. Scaffolds are coloured by  
92 genus and binned at a resolution of 30 divisions on each axis. Coloured  
93 squares within each bin are sized in proportion to the sum of individual  
94 scaffold lengths on a square-root scale, ranging from 1,005 to 771,195. The  
95 assembly has been filtered to exclude scaffolds with phylum in no-hit,  
96 Arthropoda, Ascomycota, Mollusca, undef, Streptophyta, Nematoda,  
97 Chordata, Platyhelminthes, Cnidaria, Echinodermata, Firmicutes, Annelida,  
98 Cyanobacteria, Viruses-undef, Eukaryota-undef, Actinobacteria, Tardigrada,  
99 Hemichordata, Porifera or Basidiomycota. Histograms show the distribution  
100 of scaffold length sum along each axis.

101 URL:  
102 [https://blobtoolkit.genomehubs.org/view/ACVV01/dataset/ACVV01/blob?zScale=scaleSqrt&catField=bestsumorder\\_genus&otherLimit=4&palette=user&color1=rgba%2831%2C120%2C180%2C1%29&color0=rgba%28255%2C127%2C0%2C1%29&showTotal=false&bestsumorder\\_phylum--Active=true&bestsumorder\\_phylum--Order=%2CProteobacteria&bestsumorder\\_phylum--Keys=0%2C2%2C4%2C6%2C3%2C7%2C13%2C5%2C8%2C9%2C11%2C10%2C12%2C16%2C17%2C14%2C15%2C18%2C19%2C20%2C21&bestsumorder\\_genus--Active=true&bestsumorder\\_genus--Order=Acetobacter%2CGluconobacter%2CWolbachia&color2=rgba%2851%2C160%2C44%2C1%29&plotShape=square&color3=rgba%28191%2C191%2C191%2C1%29](https://blobtoolkit.genomehubs.org/view/ACVV01/dataset/ACVV01/blob?zScale=scaleSqrt&catField=bestsumorder_genus&otherLimit=4&palette=user&color1=rgba%2831%2C120%2C180%2C1%29&color0=rgba%28255%2C127%2C0%2C1%29&showTotal=false&bestsumorder_phylum--Active=true&bestsumorder_phylum--Order=%2CProteobacteria&bestsumorder_phylum--Keys=0%2C2%2C4%2C6%2C3%2C7%2C13%2C5%2C8%2C9%2C11%2C10%2C12%2C16%2C17%2C14%2C15%2C18%2C19%2C20%2C21&bestsumorder_genus--Active=true&bestsumorder_genus--Order=Acetobacter%2CGluconobacter%2CWolbachia&color2=rgba%2851%2C160%2C44%2C1%29&plotShape=square&color3=rgba%28191%2C191%2C191%2C1%29)

### 112 Figure 3D

113 Caption: Kite-shaped blob plot of base coverage in SRR026696 against GC  
114 proportion for scaffolds in assembly ACVV01. Scaffolds are coloured by  
115 genus. Kite shapes summarise the core distribution of scaffolds. Horizontal  
116 and vertical lines represent a range spanning 2 standard deviations about  
117 the weighted mean value for each axis. The lines intersect at a point  
118 representing the weighted median value. Histograms show the distribution  
119 of scaffold length sum along each axis.

120 URL:  
121 [https://blobtoolkit.genomehubs.org/view/ACVV01/dataset/ACVV01/blob?zScale=scaleSqrt&catField=bestsumorder\\_genus&otherLimit=4&palette=u](https://blobtoolkit.genomehubs.org/view/ACVV01/dataset/ACVV01/blob?zScale=scaleSqrt&catField=bestsumorder_genus&otherLimit=4&palette=u)

123 [ser&color1=rgba%2831%2C120%2C180%2C1%29&color0=rgba%28255%2C127](#)  
124 [%2C0%2C1%29&showTotal=false&bestsumorder\\_phylum--Active=true&best](#)  
125 [sumorder\\_phylum--Order=%2CProteobacteria&bestsumorder\\_genus--Activ](#)  
126 [e=true&bestsumorder\\_genus--Order=Acetobacter%2CGluconobacter%2CW](#)  
127 [olbachia&color2=rgba%2851%2C160%2C44%2C1%29&plotShape=kite&color3](#)  
128 [=rgba%28191%2C191%2C191%2C1%29](#)

129 Figure 4A

130 Caption: Square-binned blob plot of base coverage in SRR1714990 against  
131 GC proportion for scaffolds in assembly SDAX01. Scaffolds are coloured by  
132 phylum and binned at a resolution of 30 divisions on each axis. Coloured  
133 squares within each bin are sized in proportion to the sum of individual  
134 scaffold lengths on a square-root scale, ranging from 201 to 183,673,465.  
135 Histograms show the distribution of scaffold length sum along each axis.

136 URL:

137 [https://blobtoolkit.genomehubs.org/view/Conus%20consors/dataset/SD](https://blobtoolkit.genomehubs.org/view/Conus%20consors/dataset/SDAX01/blob?staticThreshold=2800000000&nohitThreshold=2800000000)  
138 [AX01/blob?staticThreshold=2800000000&nohitThreshold=2800000000](https://blobtoolkit.genomehubs.org/view/Conus%20consors/dataset/SDAX01/blob?staticThreshold=2800000000&nohitThreshold=2800000000)

139 Figure 4B

140 Caption: Square-binned blob plot of base coverage in SRR1719763 against  
141 base coverage in SRR1712902 for scaffolds in assembly SDAX01. Scaffolds  
142 are coloured by phylum and binned at a resolution of 30 divisions on each  
143 axis. Coloured squares within each bin are sized in proportion to the sum of  
144 individual scaffold lengths on a square-root scale, ranging from 201 to  
145 33,053,995. The assembly has been filtered to exclude scaffolds with base  
146 coverage in SRR1714990 > 0.01. Histograms show the distribution of scaffold  
147 length sum along each axis.

148 URL:

149 [https://blobtoolkit.genomehubs.org/view/Conus/dataset/SDAX01/blob?S](https://blobtoolkit.genomehubs.org/view/Conus/dataset/SDAX01/blob?SRR1714990_cov--Max=0.01&SRR1719763_cov--Active=true&yField=SRR1719763_cov&xField=SRR1712902_cov&SRR1712902_cov--Active=true&staticThreshold=2800000000&nohitThreshold=2800000000)  
150 [RR1714990\\_cov--Max=0.01&SRR1719763\\_cov--Active=true&yField=SRR1719](https://blobtoolkit.genomehubs.org/view/Conus/dataset/SDAX01/blob?SRR1714990_cov--Max=0.01&SRR1719763_cov--Active=true&yField=SRR1719763_cov&xField=SRR1712902_cov&SRR1712902_cov--Active=true&staticThr)  
151 [763\\_cov&xField=SRR1712902\\_cov&SRR1712902\\_cov--Active=true&staticThr](https://blobtoolkit.genomehubs.org/view/Conus/dataset/SDAX01/blob?SRR1714990_cov--Max=0.01&SRR1719763_cov--Active=true&yField=SRR1719763_cov&xField=SRR1712902_cov&SRR1712902_cov--Active=true&staticThr)  
152 [eshold=2800000000&nohitThreshold=2800000000](https://blobtoolkit.genomehubs.org/view/Conus/dataset/SDAX01/blob?SRR1714990_cov--Max=0.01&SRR1719763_cov--Active=true&yField=SRR1719763_cov&xField=SRR1712902_cov&SRR1712902_cov--Active=true&staticThr)

153 Figure 4C

154 Caption: Square-binned blob plot of base coverage in SRR1714990 against  
155 GC proportion for scaffolds in assembly SDAX01. Scaffolds are coloured by  
156 phylum and binned at a resolution of 30 divisions on each axis. Coloured  
157 squares within each bin are sized in proportion to the sum of individual  
158 scaffold lengths on a square-root scale, ranging from 201 to 21,767,682. The  
159 assembly has been filtered to exclude scaffolds with phylum matches

160 no-hit. Histograms show the distribution of scaffold length sum along each  
161 axis.

162 URL:

163 <https://blobtoolkit.genomehubs.org/view/Conus%20consors/dataset/SD>  
164 [AX01/blob?bestsumorder\\_phylum--Keys=6&staticThreshold=2800000000](https://blobtoolkit.genomehubs.org/view/Conus%20consors/dataset/SD)

165 Figure 4D

166 Caption: Blob plot of base coverage in SRR1714990 against GC proportion  
167 for scaffolds in assembly SDAX01. Scaffolds are coloured by phylum. Circles  
168 are sized in proportion to scaffold length on a square-root scale, ranging  
169 from 201 to 148,256. The assembly has been filtered to exclude scaffolds  
170 with phylum matches no-hit. Histograms show the distribution of scaffold  
171 length sum along each axis.

172 URL:

173 <https://blobtoolkit.genomehubs.org/view/Conus%20consors/dataset/SD>  
174 [AX01/blob?bestsumorder\\_phylum--Keys=6&plotShape=circle&plotGraphic](https://blobtoolkit.genomehubs.org/view/Conus%20consors/dataset/SD)  
175 [s=svg&staticThreshold=2800000000](https://blobtoolkit.genomehubs.org/view/Conus%20consors/dataset/SD)

176 Figure 4E

177 Caption: Kite-shaped blob plot of base coverage in SRR1714990 against GC  
178 proportion for scaffolds in assembly SDAX01. Scaffolds are coloured by  
179 phylum. Kite shapes summarise the core distribution of scaffolds.  
180 Horizontal and vertical lines represent a range spanning 2 standard  
181 deviations about the weighted mean value for each axis. The lines intersect  
182 at a point representing the weighted median value. Histograms show the  
183 distribution of scaffold length sum along each axis. The assembly has been  
184 filtered to exclude scaffolds with phylum matches no-hit.

185 URL:

186 <https://blobtoolkit.genomehubs.org/view/Conus%20consors/dataset/SD>  
187 [AX01/blob?bestsumorder\\_phylum--Keys=6&plotShape=kite&staticThreshol](https://blobtoolkit.genomehubs.org/view/Conus%20consors/dataset/SD)  
188 [d=2800000000](https://blobtoolkit.genomehubs.org/view/Conus%20consors/dataset/SD)

189 Figure 5A

190 Caption: Blob plot of base coverage in SRR6918124 against GC proportion  
191 for scaffolds in assembly PTEZ01. Scaffolds are coloured by phylum. Circles  
192 are sized in proportion to scaffold length on a square-root scale, ranging  
193 from 1,000 to 14,341,824. Histograms show the distribution of scaffold  
194 length sum along each axis.

195 URL:

196 <https://blobtoolkit.genomehubs.org/view/Aves/dataset/PTEZ01/blob?plo>  
197 [tShape=circle&plotResolution=42&yField=SRR6918124\\_cov&SRR6918124\\_co](https://blobtoolkit.genomehubs.org/view/Aves/dataset/PTEZ01/blob?plo)

198 [v--Active=true&palette=user&color1=rgba%28191%2C191%2C191%2C1%29&c](https://blobtoolkit.genomehubs.org/view/Aves/dataset/PTEZ01/blob?plotShape=circle&plotResolution=42&yField=SRR6918124_cov&SRR6918124_cov--Active=true&palette=user&color1=rgba%28191%2C191%2C191%2C1%29&color2=rgba%28255%2C127%2C0%2C1%29&color7=rgba%28178%2C223%2C138%2C1%29&color0=rgba%2831%2C120%2C180%2C1%29#Settings)  
199 [olor2=rgba%28255%2C127%2C0%2C1%29&color7=rgba%28178%2C223%2C13](https://blobtoolkit.genomehubs.org/view/Aves/dataset/PTEZ01/blob?plotShape=circle&plotResolution=42&yField=SRR6918124_cov&SRR6918124_cov--Active=true&palette=user&color1=rgba%28191%2C191%2C191%2C1%29&color2=rgba%28255%2C127%2C0%2C1%29&color7=rgba%28178%2C223%2C138%2C1%29&color0=rgba%2831%2C120%2C180%2C1%29#Settings)  
200 [8%2C1%29&color0=rgba%2831%2C120%2C180%2C1%29#Settings](https://blobtoolkit.genomehubs.org/view/Aves/dataset/PTEZ01/blob?plotShape=circle&plotResolution=42&yField=SRR6918124_cov&SRR6918124_cov--Active=true&palette=user&color1=rgba%28191%2C191%2C191%2C1%29&color2=rgba%28255%2C127%2C0%2C1%29&color7=rgba%28178%2C223%2C138%2C1%29&color0=rgba%2831%2C120%2C180%2C1%29#Settings)

## 201 Figure 5B

202 Caption: Blob plot of base coverage in SRR6918124 against GC proportion  
203 for scaffolds in assembly PTEZ01. Scaffolds are coloured by family. Circles  
204 are sized in proportion to scaffold length on a square-root scale, ranging  
205 from 1,000 to 613,323. The assembly has been filtered to exclude: scaffolds  
206 with base coverage in SRR6918124 > 2; or family not in no-hit, Physeteridae,  
207 Tinamidae, Odontophoridae or Sarcocystidae. Histograms show the  
208 distribution of scaffold length sum along each axis.

209 URL:

210 [https://blobtoolkit.genomehubs.org/view/Aves/dataset/PTEZ01/blob?plo](https://blobtoolkit.genomehubs.org/view/Aves/dataset/PTEZ01/blob?plotShape=circle&plotResolution=42&yField=SRR6918124_cov&SRR6918124_cov--Active=true&palette=user&color1=rgba%28191%2C191%2C191%2C1%29&color2=rgba%28255%2C127%2C0%2C1%29&color7=rgba%28178%2C223%2C138%2C1%29&bestsumorder_family--Active=true&catField=bestsumorder_family&color4=rgba%2851%2C160%2C44%2C1%29&color3=rgba%2851%2C160%2C44%2C1%29&color6=rgba%2831%2C120%2C180%2C1%29&color0=rgba%28166%2C206%2C227%2C1%29&bestsumorder_family--Keys=13%2C5%2C6%2C9%2C10&bestsumorder_family--Inv=true&SRR6918124_cov--Max=2)  
211 [tShape=circle&plotResolution=42&yField=SRR6918124\\_cov&SRR6918124\\_co](https://blobtoolkit.genomehubs.org/view/Aves/dataset/PTEZ01/blob?plotShape=circle&plotResolution=42&yField=SRR6918124_cov&SRR6918124_cov--Active=true&palette=user&color1=rgba%28191%2C191%2C191%2C1%29&color2=rgba%28255%2C127%2C0%2C1%29&color7=rgba%28178%2C223%2C138%2C1%29&bestsumorder_family--Active=true&catField=bestsumorder_fa)  
212 [v--Active=true&palette=user&color1=rgba%28191%2C191%2C191%2C1%29&c](https://blobtoolkit.genomehubs.org/view/Aves/dataset/PTEZ01/blob?plotShape=circle&plotResolution=42&yField=SRR6918124_cov&SRR6918124_cov--Active=true&palette=user&color1=rgba%28191%2C191%2C191%2C1%29&color2=rgba%28255%2C127%2C0%2C1%29&color7=rgba%28178%2C223%2C138%2C1%29&bestsumorder_family--Active=true&catField=bestsumorder_fa)  
213 [olor2=rgba%28255%2C127%2C0%2C1%29&color7=rgba%28178%2C223%2C13](https://blobtoolkit.genomehubs.org/view/Aves/dataset/PTEZ01/blob?plotShape=circle&plotResolution=42&yField=SRR6918124_cov&SRR6918124_cov--Active=true&palette=user&color1=rgba%28191%2C191%2C191%2C1%29&color2=rgba%28255%2C127%2C0%2C1%29&color7=rgba%28178%2C223%2C138%2C1%29&bestsumorder_family--Active=true&catField=bestsumorder_fa)  
214 [8%2C1%29&bestsumorder\\_family--Active=true&catField=bestsumorder\\_fa](https://blobtoolkit.genomehubs.org/view/Aves/dataset/PTEZ01/blob?plotShape=circle&plotResolution=42&yField=SRR6918124_cov&SRR6918124_cov--Active=true&palette=user&color1=rgba%28191%2C191%2C191%2C1%29&color2=rgba%28255%2C127%2C0%2C1%29&color7=rgba%28178%2C223%2C138%2C1%29&bestsumorder_family--Active=true&catField=bestsumorder_fa)  
215 [mily&color4=rgba%2851%2C160%2C44%2C1%29&color3=rgba%2851%2C160%](https://blobtoolkit.genomehubs.org/view/Aves/dataset/PTEZ01/blob?plotShape=circle&plotResolution=42&yField=SRR6918124_cov&SRR6918124_cov--Active=true&palette=user&color1=rgba%28191%2C191%2C191%2C1%29&color2=rgba%28255%2C127%2C0%2C1%29&color7=rgba%28178%2C223%2C138%2C1%29&bestsumorder_family--Active=true&catField=bestsumorder_fa)  
216 [2C44%2C1%29&color3=rgba%2851%2C160%](https://blobtoolkit.genomehubs.org/view/Aves/dataset/PTEZ01/blob?plotShape=circle&plotResolution=42&yField=SRR6918124_cov&SRR6918124_cov--Active=true&palette=user&color1=rgba%28191%2C191%2C191%2C1%29&color2=rgba%28255%2C127%2C0%2C1%29&color7=rgba%28178%2C223%2C138%2C1%29&bestsumorder_family--Active=true&catField=bestsumorder_fa)  
217 [2C44%2C1%29&color6=rgba%2831%2C120%2C180%2C1%29&color0=rgba%28](https://blobtoolkit.genomehubs.org/view/Aves/dataset/PTEZ01/blob?plotShape=circle&plotResolution=42&yField=SRR6918124_cov&SRR6918124_cov--Active=true&palette=user&color1=rgba%28191%2C191%2C191%2C1%29&color2=rgba%28255%2C127%2C0%2C1%29&color7=rgba%28178%2C223%2C138%2C1%29&bestsumorder_family--Active=true&catField=bestsumorder_fa)  
218 [166%2C206%2C227%2C1%29&bestsumorder\\_family--Keys=13%2C5%2C6%2C](https://blobtoolkit.genomehubs.org/view/Aves/dataset/PTEZ01/blob?plotShape=circle&plotResolution=42&yField=SRR6918124_cov&SRR6918124_cov--Active=true&palette=user&color1=rgba%28191%2C191%2C191%2C1%29&color2=rgba%28255%2C127%2C0%2C1%29&color7=rgba%28178%2C223%2C138%2C1%29&bestsumorder_family--Active=true&catField=bestsumorder_fa)

## 219 Figure 5C

220 Caption: Square-binned blob plot of base coverage in SRR6918124 against  
221 GC proportion for scaffolds in assembly PTEZ01. Scaffolds are coloured by  
222 phylum and binned at a resolution of 42 divisions on each axis. Coloured  
223 squares within each bin are sized in proportion to the sum of individual  
224 scaffold lengths on a square-root scale, ranging from 1,000 to 172,797,723.  
225 Histograms show the distribution of scaffold length sum along each axis.

226 URL:

227 [https://blobtoolkit.genomehubs.org/view/Aves/dataset/PTEZ01/blob?plotResol](https://blobtoolkit.genomehubs.org/view/Aves/dataset/PTEZ01/blob?plotResolution=42&yField=SRR6918124_cov&SRR6918124_cov--Active=true&palette=user&color1=rgba%28191%2C191%2C191%2C1%29&color2=rgba%28255%2C127%2C0%2C1%29&color7=rgba%28178%2C223%2C138%2C1%29&color0=rgba%2831%2C120%2C180%2C1%29#Lists)  
228 [ution=42&yField=SRR6918124\\_cov&SRR6918124\\_cov--Active=true&palette=user&c](https://blobtoolkit.genomehubs.org/view/Aves/dataset/PTEZ01/blob?plotResolution=42&yField=SRR6918124_cov&SRR6918124_cov--Active=true&palette=user&color1=rgba%28191%2C191%2C191%2C1%29&color2=rgba%28255%2C127%2C0%2C1%29&color7=rgba%28178%2C223%2C138%2C1%29&color0=rgba%2831%2C120%2C180%2C1%29#Lists)  
229 [olor1=rgba%28191%2C191%2C191%2C1%29&color2=rgba%28255%2C127%2C0%2C1%](https://blobtoolkit.genomehubs.org/view/Aves/dataset/PTEZ01/blob?plotResolution=42&yField=SRR6918124_cov&SRR6918124_cov--Active=true&palette=user&color1=rgba%28191%2C191%2C191%2C1%29&color2=rgba%28255%2C127%2C0%2C1%29&color7=rgba%28178%2C223%2C138%2C1%29&color0=rgba%2831%2C120%2C180%2C1%29#Lists)  
230 [29&color7=rgba%28178%2C223%2C138%2C1%29&color0=rgba%2831%2C120%](https://blobtoolkit.genomehubs.org/view/Aves/dataset/PTEZ01/blob?plotResolution=42&yField=SRR6918124_cov&SRR6918124_cov--Active=true&palette=user&color1=rgba%28191%2C191%2C191%2C1%29&color2=rgba%28255%2C127%2C0%2C1%29&color7=rgba%28178%2C223%2C138%2C1%29&color0=rgba%2831%2C120%2C180%2C1%29#Lists)  
231 [%2C1%29&#Lists](https://blobtoolkit.genomehubs.org/view/Aves/dataset/PTEZ01/blob?plotResolution=42&yField=SRR6918124_cov&SRR6918124_cov--Active=true&palette=user&color1=rgba%28191%2C191%2C191%2C1%29&color2=rgba%28255%2C127%2C0%2C1%29&color7=rgba%28178%2C223%2C138%2C1%29&color0=rgba%2831%2C120%2C180%2C1%29#Lists)

232 Selection: To highlight the set of scaffolds with BUSCO annotations in any  
233 of the reference gene sets, upload File S3 to the above URL then  
234 deactivate the selection using the Filters menu.

235

## Table reproduction

236 Tables were produced using code available in the **INSDC-pipeline** repository or  
237 obtained directly from the Viewer, using the commands and URLs presented  
238 below.

### 239 Table 1

240 Current values for the numbers of analysed and available assemblies  
241 presented in Table 1 can be obtained using the **INSDC-pipeline** script  
242 `available_assemblies.py` and a local copy of the NCBI taxonomy  
243 `new_taxdump` directory:

```
244 scripts/available_assemblies.py --taxdump  
245 /path/to/new_taxdump
```

### 246 Table 2

247 Values in Table 2 were obtained from the BUSCO view of the unfiltered  
248 PTEZ01 dataset  
249 (<https://blobtoolkit.genomehubs.org/view/Aves/dataset/PTEZ01/busco>).

250 Numbers in parentheses were obtained by subtracting the BUSCO scores  
251 with a minimum coverage filter of 2 applied to the base coverage in read  
252 set SRR6918124

253 ([https://blobtoolkit.genomehubs.org/view/Aves/dataset/PTEZ01/busco?  
254 SRR6918124\\_cov--Active=true&SRR6918124\\_cov--Min=2](https://blobtoolkit.genomehubs.org/view/Aves/dataset/PTEZ01/busco?SRR6918124_cov--Active=true&SRR6918124_cov--Min=2)) from the unfiltered  
255 values.

### 256 Table 3

257 Values in Table 3 were obtained from  
258 <https://blobtoolkit.genomehubs.org/view/Chordata> by using the  
259 *Customise table* option and adding *Apicomplexa span* and *Chordata span* to  
260 the displayed assembly statistics.