

Supplementary Figures for

Estimating chromosome sizes from karyotype images
enables validation of *de novo* genome assemblies.

Arne Ludwig^{*,1,2}, Alexandr Dibrov^{*,1,2}, Gene Myers^{1,2}, Martin Pippel^{†,1,2}

January 6, 2023

*Equal contribution

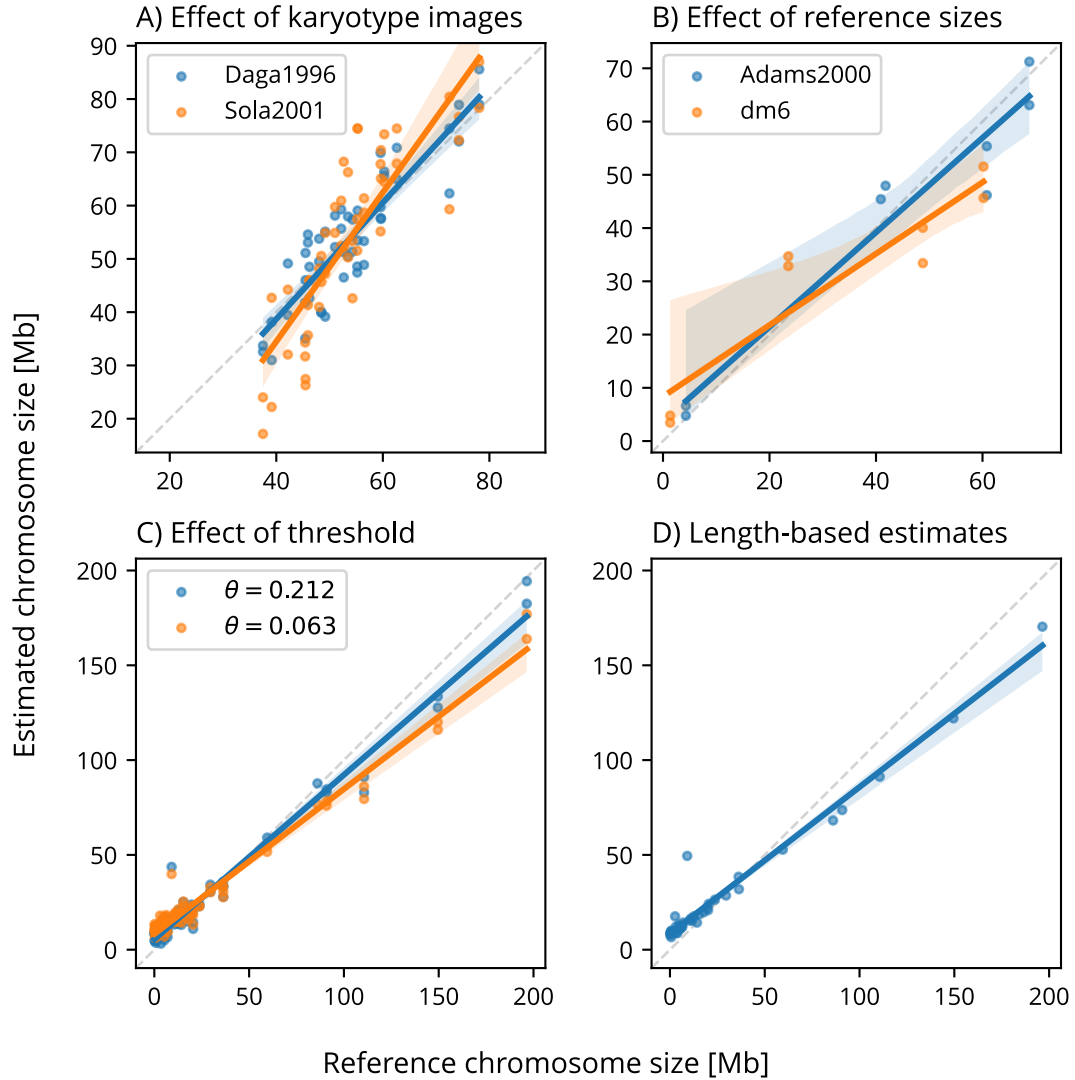
†To whom correspondence should be addressed:

Martin Pippel
Max Planck Institute of Molecular Cell Biology and Genetics
Pfotenhauerstr. 108
01307 Dresden, Germany
Tel: +49 351 210-1972
Mail: pippel@mpi-cbg.de

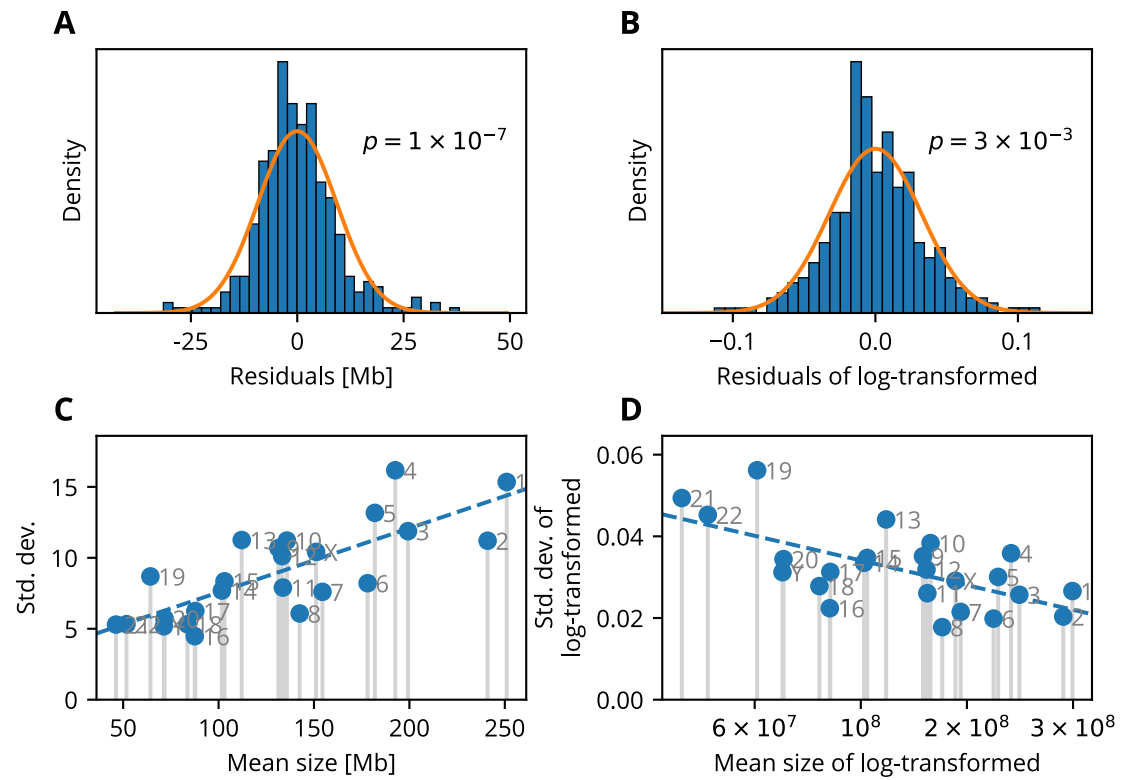
¹Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstr. 108, 01307 Dresden, Germany

²Center for Systems Biology Dresden, Pfotenhauerstr. 108, 01307 Dresden, Germany

Supplementary Figures



Supplementary Figure 1: Examples for the effects of varying quality of analysis inputs. **A)** Estimate results for *D. rerio* with karyotype images from Daga1996 [2] and Sola2001 [4]. **B)** Estimate results for *D. melanogaster* with reference sizes from Adams2000 [1] and the current reference assembly dm6 (GCA_000001215.4). **C)** Estimate results for *G. gallus* with thresholds $\theta = 0.212$ and $\theta = 0.063$. **D)** Estimates based on chromosome lengths found in [3].



Supplementary Figure 2: Analysis of residual distribution using the HUMAN8 dataset. The p -values stated in the plots are derived from the Shapiro-Wilk test. **A)** Distribution of residuals of plain chromosome size estimates. **B)** Distribution of residuals of log-transformed chromosome size estimates. **C)** Standard deviation over mean of plain chromosome size estimates. The dashed regression line highlights a trend that larger chromosomes have a higher standard deviation. **D)** Standard deviation over mean of log-transformed chromosome size estimates. The dashed regression line highlights a trend that smaller chromosomes have a higher standard deviation after log-transform.

References

- [1] Mark D. Adams et al. "The Genome Sequence of *Drosophila melanogaster*." In: *Science* 287.5461 (Mar. 2000), pp. 2185–2195. DOI: 10.1126/science.287.5461.2185. eprint: <https://www.science.org/doi/pdf/10.1126/science.287.5461.2185>. URL: <https://www.science.org/doi/abs/10.1126/science.287.5461.2185>.
- [2] Rafael R. Daga, Guillermo Thode, and Angel Amores. "Chromosome complement, C-banding, Ag-NOR and replication banding in the zebrafish *Danio reio*."

In: *Chromosome Research* 4.1 (1996), pp. 29–32. ISSN: 1573-6849. URL: <https://doi.org/10.1007/BF02254941>.

- [3] Alberto Juan Solari. "Ultrastructure of the synaptic autosomes and the ZW bivalent in chicken oocytes." In: *Chromosoma* 64.2 (1977), pp. 155–165. ISSN: 1432-0886. DOI: 10.1007/BF00327055. URL: <https://doi.org/10.1007/BF00327055>.
- [4] L. Sola and E. Gornung. "Classical and molecular cytogenetics of the zebrafish, *Danio rerio* (Cyprinidae, Cypriniformes): an overview." In: *Genetica* 111.1 (2001), pp. 397–412. ISSN: 1573-6857. URL: <https://doi.org/10.1023/A:1013776323077>.