This dataset contains all original files used for calculating pressure and forces, as well as the final results.

The following descriptions are organized based on the folder tree view.

Raw data files:

- Body depth profiles – lateral view images of the fish, the manually-digitized outlines of the fish's body, and a ReadMe explaining how the pixel to mm calibration factors were set (and handling of a depth-of-field error for trout4).

- Velocity data from PIV – velocity vector fields derived from the video clips. The fields have been cropped to a 18x18 cm window based on the convergence analysis described in the associated paper. These files have been formatted from native LaVision DaVis 8.2.2 output for compatibility with pressure and force calculation scripts. While PIV was conducted using every video frame from the 1000 fps film, our validations (Lucas et al. 2017, PLoS ONE) indicated that pressure calculation only requires fields at a time-step of 0.01s. As such, only every 10th vector field (those used in pressure calculations) has been uploaded.

- Ventral view outlines – manually-digitized outlines of the fish for every 10th video frame. These are used to generate 1) masks of the fish for queen2, 2) the force calculation boundaries, and 3) silhouettes of the fish in figures. Midlines are automatically extracted from these outlines, too.

- Video clips – Clips from original research videos filmed at 1000 fps. These have been clipped to contain ~1.5 tailbeat cycles. See associated manuscript for explanation of where force calculations are conducted.

Results:

Contained within "Calculated pressures and forces." Results are organized by video sequence.

Note: Force calculation scripts can be found at
https://github.com/kelseynlucas/Forces-on-carangiform-swimmers

For dat files, each column represents a frame of video. Each row represents a point on the force calculation boundary. Coordinates of these points can be found in the xvalsBound and yvalsBound files (see below).

Forces were extracted starting from the 3rd pressure field in each sequence and ending after a complete tailbeat cycle. Because queen2 includes temporal smoothing (see Dabiri et al. 2014,

Journal of Experimental Biology; Lucas et al. 2017, PLoS ONE), the first few time-steps of the sequence are excluded from force calculation to account for end behavior of the splines.

- distOnMidline – a list of how far along the body, where 0 is the nose and 1 is the tip of the caudal fin, each force calculation boundary point is. Used to assign force data to different body segments. One file per sequence.

- Midlines – coordinates of midlines extracted automatically from the fish outlines. These are smoothed and then used to extract kinematics information. One file per sequence.

- Pressures – surface pressure at each force calculation boundary point (used to calculate forces). One file per sequence.

- Queen2 file stack – pressure field for every 10^{th} frame of video (numbered by the frame of the original video). See queen2 ReadMe for explanation of data format.

- xForcesAll – x-component of force vectors. Note, since fish swim to the left side of the screen, negative x-components represent contributions to thrust. One file per sequence.

- yForcesAll – y-component of force vectors. One file per sequence.

- xUnitNormals – x-component of the outward-facing normal vector describing the surface of the force calculation boundary. Used to calculate forces and to assign forces to thrust and drag directions. One file per sequence.

- yUnitNormals – y-component of the outward-facing normal vector describing the surface of the force calculation boundary. Used to calculate forces and to assign forces to the fish's left and right directions. One file per sequence.

- xvalsBound – x-coordinates of the force calculation boundary points. Used to query surface pressure values for force calculation. One file per sequence.

- yvalsBound – y-coordinates of the force calculation boundary points. Used to query surface pressure values for force calculation. One file per sequence.

Statistics:

The provided data files are aggregations of the above data, formatted for statistical analyses. The R code used to perform statistical analyses is also provided here in three formats: Rmd (R-

markdown; original file, including the analysis results), and html and pdf (for viewing without an R installation).