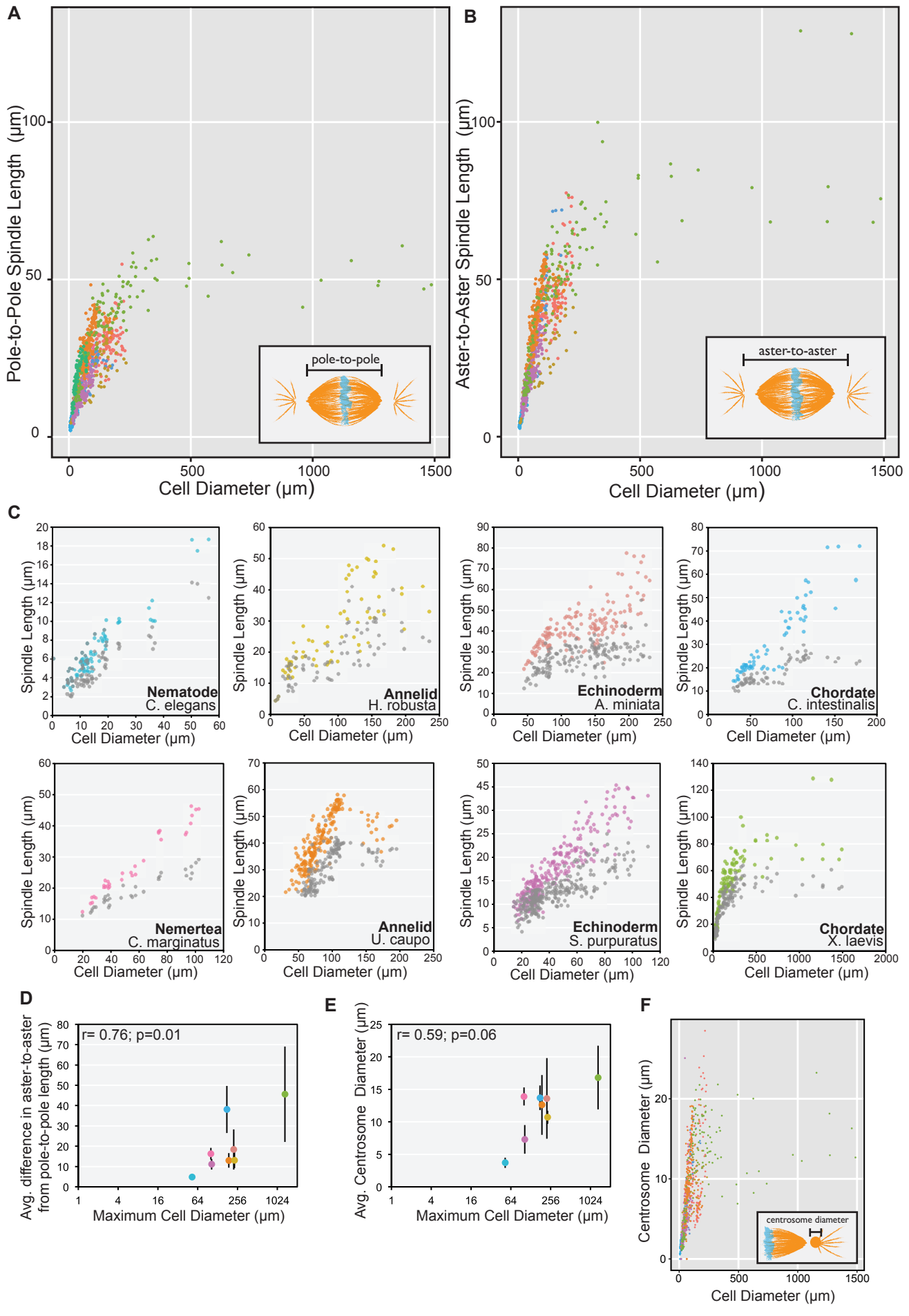
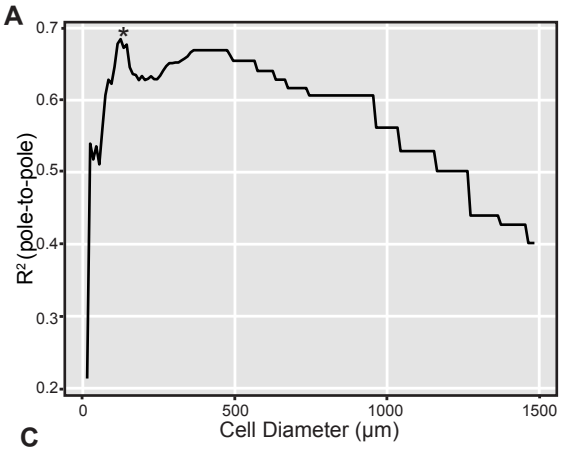
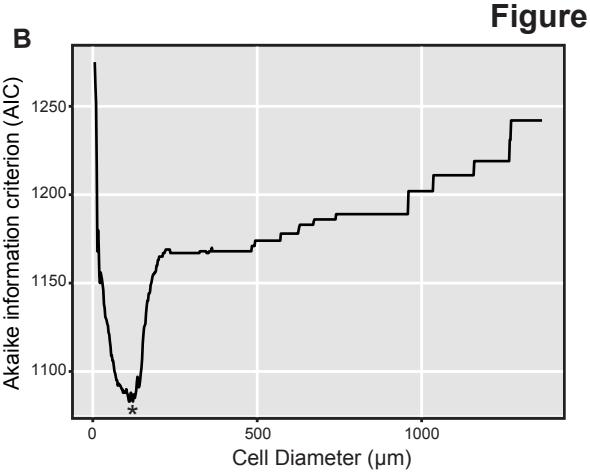
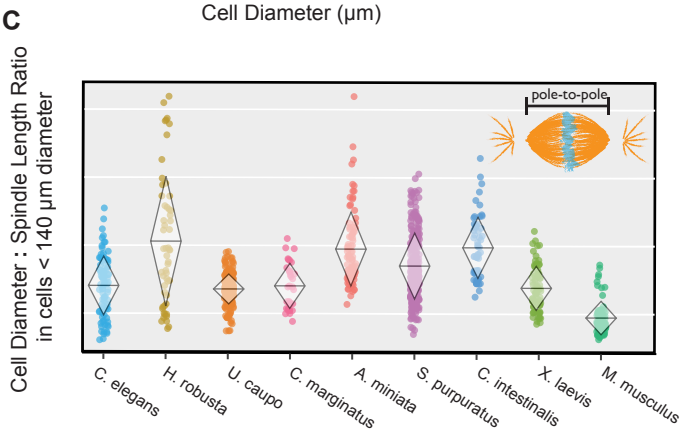
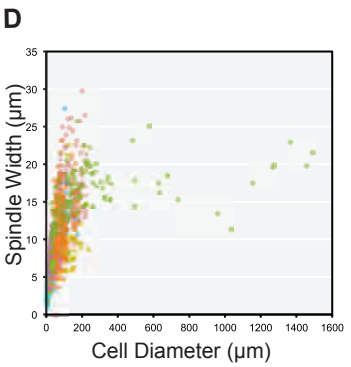

A
Pole-to-Pole Spindle Length (µm)
Cell Diameter (µm)
pole-to-pole
B
Aster-to-Aster Spindle Length (µm)
Cell Diameter (µm)
aster-to-aster
C
Spindle Length (µm)
Cell Diameter (µm)
Nematode
C. elegans
Annelid
H. robusta
Echinoderm
A. miniata
Chordate
C. intestinalis
Nemertea
C. marginatus
Annelid
U. caupo
Echinoderm
S. purpuratus
Chordate
X. laevis
D
Avg. difference in aster-to-aster from pole-to-pole length (µm)
r= 0.76; p=0.01
Maximum Cell Diameter (µm)
E
Avg. Centrosome Diameter (µm)
r= 0.59; p=0.06
Maximum Cell Diameter (µm)
F
Centrosome Diameter (µm)
centrosome diameter
Cell Diameter (µm)

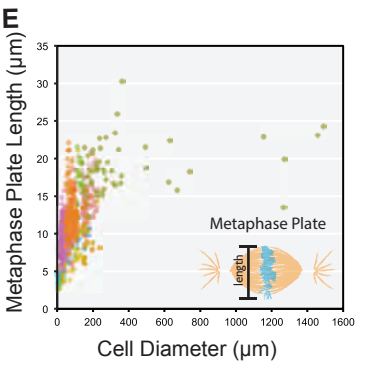
Figure S2

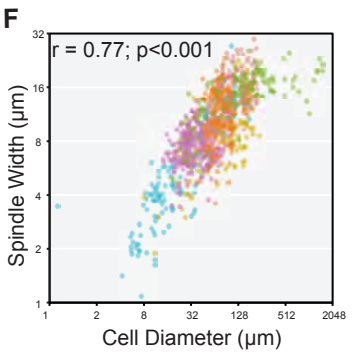

A
R^2 (pole-to-pole)
Cell Diameter (µm)

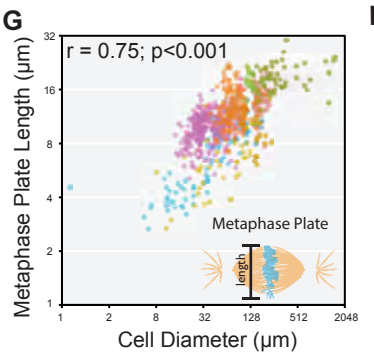

B
Akaike information criterion (AIC)
Cell Diameter (µm)


C
Cell Diameter : Spindle Length Ratio in cells < 140 µm diameter
pole-to-pole
C. elegans
H. robusta
U. caupo
C. marginatus
A. miniata
S. purpuratus
C. intestinalis
X. laevis
M. musculus


D
Spindle Width (µm)
Cell Diameter (µm)


E
Metaphase Plate Length (µm)
Metaphase Plate
length
Cell Diameter (µm)


F
r = 0.77; p<0.001
Spindle Width (µm)
Cell Diameter (µm)


G
r = 0.75; p<0.001
Metaphase Plate Length (µm)
Metaphase Plate
length
Cell Diameter (µm)

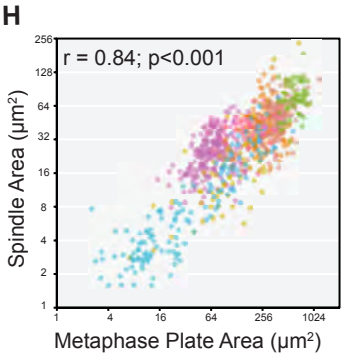

H
r = 0.84; p<0.001
Spindle Area (µm²)
Metaphase Plate Area (µm²)

Figure S3

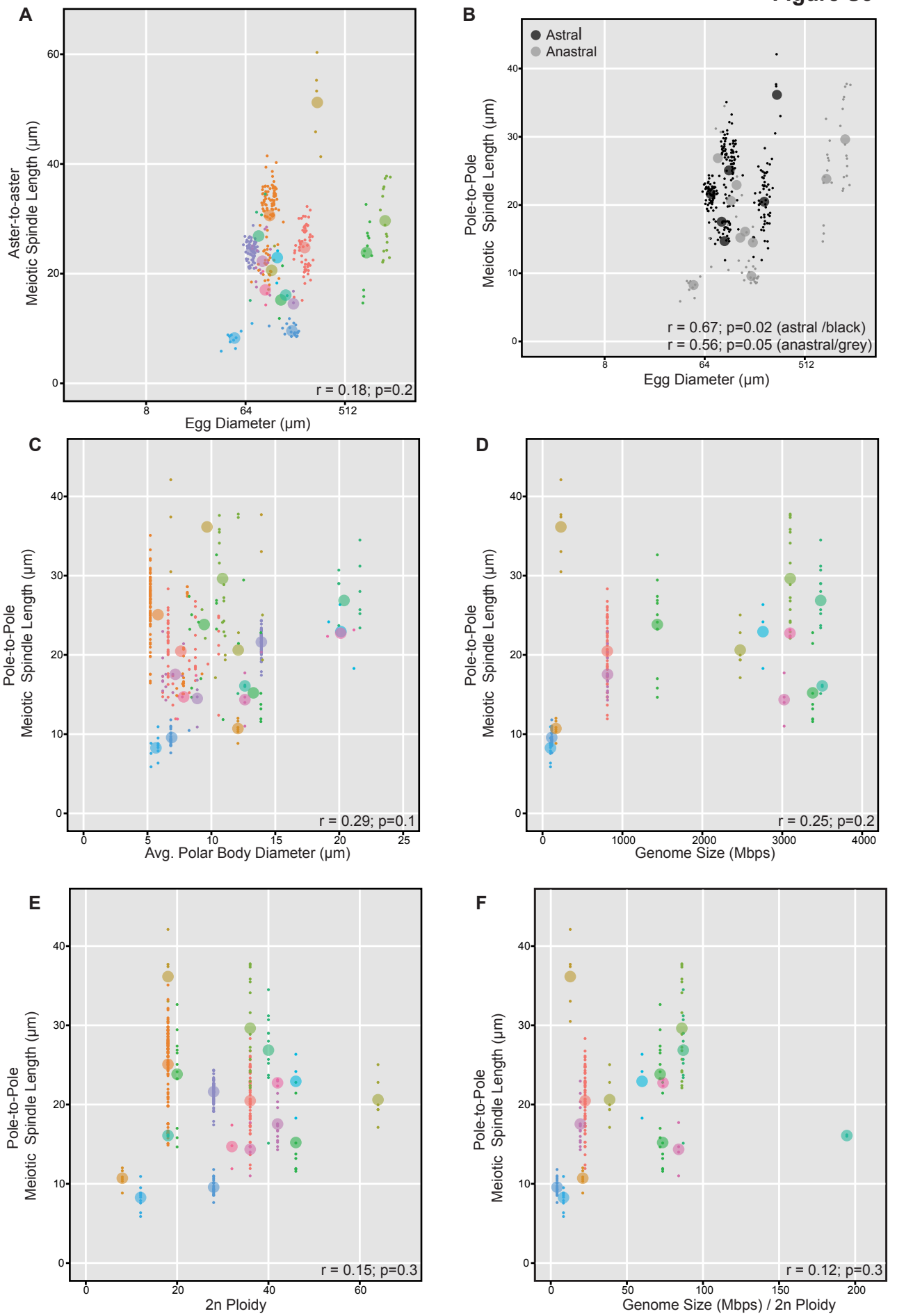
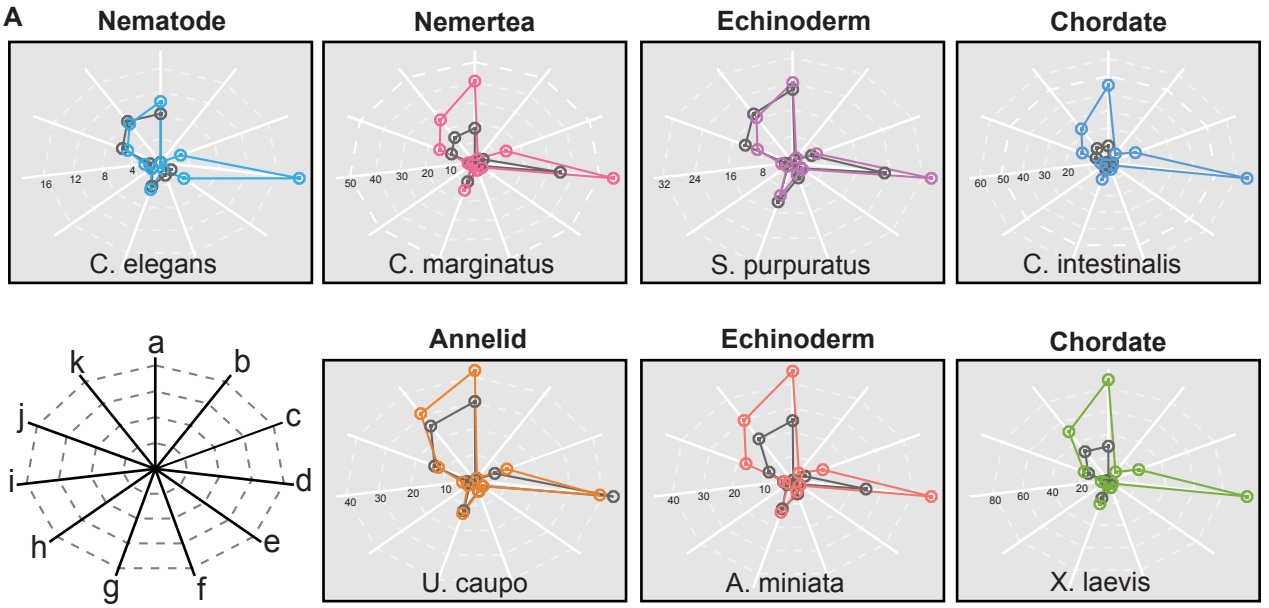

A
Aster-to-aster Meiotic Spindle Length (µm)
Egg Diameter (µm)
r = 0.18; p=0.2
B
Astral
Anastral
Pole-to-Pole Meiotic Spindle Length (µm)
Egg Diameter (µm)
r = 0.67; p=0.02 (astral /black)
r = 0.56; p=0.05 (anastral/grey)
C
Pole-to-Pole Meiotic Spindle Length (µm)
Avg. Polar Body Diameter (µm)
r = 0.29; p=0.1
D
Pole-to-Pole Meiotic Spindle Length (µm)
Genome Size (Mbps)
r = 0.25; p=0.2
E
Pole-to-Pole Meiotic Spindle Length (µm)
2n Ploidy
r = 0.15; p=0.3
F
Pole-to-Pole Meiotic Spindle Length (µm)
Genome Size (Mbps) / 2n Ploidy
r = 0.12; p=0.3

Figure S4


A
Nematode
Nemertea
Echinoderm
Chordate
16 12 8 4
50 40 30 20 10
32 24 16 8
60 50 40 30 20
C. elegans
C. marginatus
S. purpuratus
C. intestinalis
a
b
c
d
e
f
g
h
i
j
k
Annelid
Echinoderm
Chordate
40 30 20 10
40 30 20 10
80 60 40 20
U. caupo
A. miniata
X. laevis

Supplemental Figure Legends

Figure S1. (A) Mitotic pole-to-pole spindle length versus cell diameter in early embryonic divisions of various metazoan species. (B) Mitotic aster-to-aster spindle length versus cell diameter in early embryonic divisions of various metazoan species. (C) Individual species plots of mitotic pole-to-pole (gray) and aster-to-aster (color) spindle length versus cell diameter in early embryonic divisions. Plots display measurements for all cells/embryos analyzed where individual data points represent a single spindle measurement and different colors represent different species. (D) Average difference between the aster-to-aster and pole-to-pole spindle length versus cell diameter (log_2 scale) in the first and second embryonic divisions of various metazoan species, r= 0.76, p=0.01. (E) Average centrosome diameter versus cell diameter (log_2 scale) in the first and second embryonic divisions of various metazoan species, r= 0.59, p=0.06. Error bars represent standard deviation of the mean. (F) Centrosome diameter versus cell diameter in early embryonic divisions of various metazoan species showing measurements for all cells/embryos analyzed. Different colors represent different species as indicated in Figure 1A key.

Figure S2. (A) Plot of R^2 values versus a range of cell diameter breakpoints revealing the linear scaling regime (asterisk) for pole-to-pole length measurements of all mitotic spindles. A higher R^2 value represents a better fit to a linear regression model. (B) Plot of the Akaike information criterion (AIC) versus a range of cell diameter breakpoints for two-segment piecewise regression model (mitotic spindle pole-to-pole lengths). The AIC value represents a penalized log-likelihood estimate of a model. It contains a penalty

term for each additional parameter that accounts for the risk of overfitting. A lower AIC value represents a better model (both in terms of the fraction of variation explained and a number of parameters (asterisk = breakpoint in two-segment model). (C) Cell diameter : pole-to-pole spindle length ratio of cells with cell diameter < 140 μm for individual species. Overlaid diamonds indicate mean cell diameter : spindle length ratio (center horizontal line) and standard deviation (height of vertices) for each species. (D) Mitotic spindle width versus cell diameter in early embryonic divisions of various metazoan species for all cells analyzed. (E) Mitotic metaphase plate length versus cell diameter in early embryonic divisions of various metazoan species for all cells analyzed. (F) Mitotic spindle width on a $\log_2$ scale versus cell diameter on a $\log_2$ scale in early embryonic divisions of various metazoan species for all cells analyzed, $r= 0.77$, $p<0.001$. (G) Mitotic metaphase plate length on a $\log_2$ scale versus cell diameter on a $\log_2$ scale in early embryonic divisions of various metazoan species for all cells analyzed, $r= 0.75$, $p<0.001$. (H) Spindle area (pole-to-pole spindle length x spindle width) on a $\log_2$ scale versus metaphase plate area (metaphase plate length x width) on a $\log_2$ scale in early embryonic divisions of various metazoan species, $r= 0.84$, $p<0.001$. Individual data points represent a single spindle measurement and different colors represent different species as indicated in Figure 1A key.

Figure S3. (A) Female meiotic aster-to-aster spindle length versus cell diameter ($\log_2$ scale) from eggs of various metazoan species, $r=0.18$, $p=0.2$. (B) Female meiotic pole-to-pole spindle length for astral (black, $r=0.67$, $p=0.02$) and anastral (grey, $r= 0.56$, $p=0.05$) meiotic spindles versus cell diameter ($\log_2$ scale) from eggs of various

metazoan species. (C) Female meiotic pole-to-pole spindle length versus average polar body diameter for various metazoan species, r= 0.29, p=0.1. (D) Female meiotic pole-to-pole spindle length versus genome size for various metazoan species, r= 0.25, p=0.2. (E) Female meiotic pole-to-pole spindle length versus diploid chromosome number for various metazoan species, r= 0.15 p=0.3 (F) Female meiotic pole-to-pole spindle length versus genome size normalized to diploid chromosome number for various metazoan species, r= 0.12, p=0.3. All plots display measurements for all female meiotic spindles analyzed where individual data points represent a single spindle measurements and different colors represent different species as indicated in Figure 3C. Larger data points represent averages of all measurements for individual species.

Figure S4. (A) Spider plots of measurement averages (in μm) of spindle parameters from metaphase mitotic spindles (color) from 1- to 4-cell embryos and female meiotic spindles (grey) for each species. Each axis represents a different spindle parameter: a= aster-to-aster spindle length, b= distance between centrosome aster-to-spindle pole, c= centrosome aster diameter, d= astral microtubule array diameter, e= ratio of metaphase plate length : width, f= metaphase plate width, g= metaphase plate length, h= ratio of pole-to-pole spindle length : width, i= ratio of aster-to-aster spindle length : width, j= spindle width, k= pole-to-pole spindle length.

Supplemental Experimental Procedures

Spawning, in vitro fertilization, and embryo collection

Wild animals were collected from Bodega Bay, CA (*A. miniata, S. purpuratus, U. caupo, M. californianus*), Coos Bay, OR (*C. cf. marginatus),* or obtained from Marine Research and Education Products (M-REP), Carlsbad, CA (*C. intestinalis*). *X. laevis, C. elegans,* and *H. robusta* were obtained from laboratory colonies. We followed established methods to collect eggs and sperm, fertilize in vitro, and culture developing embryos for each of the following organisms: *Xenopus laevis* [S1], *Ciona intestinalis* [S2], *Urechis caupo* [S3, S4], *Asterina miniata* [S5], *Strongylocentrotus purpuratus* [S6], *Mytilus californianus* [S7] and *Cerebratulus cf. marginatus* [S8]. *Cerebratulus* used here conforms to the description of *C. marginatus* [S9], but is likely an undescribed species. All eggs were collected by induced spawning, with the exception of *C. marginatus* from which eggs were obtained by perforating the body wall of gravis animals and *S. purpuratus* from which meiotic eggs were obtained by dissecting intact ovaries from gravid adults. *Caenorhabditis elegans* and *Helobdella robusta* self-fertilize; eggs and fertilized embryos were collected using previously described methods for *C. elegans* [S10] and *H. robusta* [S11].

Immunofluorescence and Imaging

Embryos were collected and fixed at various time points after fertilization. Embryos from each organism required different treatments to remove membrane or jelly coats for efficient fixation and antibody staining. Briefly, the jelly coat of *X. laevis* embryos were removed by treatment with 2% cysteine (Sigma) in 1/3X MMR [S1]. The fertilization membrane of *U. caupo* embryos was removed by treatment with 0.54 M Glucose in seawater following previously described methods [S4]. The fertilization membrane of *A.*

miniata and *S. purpuratus* embryos was removed by treatment with 5 μM PABA (Sigma) in filtered seawater. The vitelline envelope of *H. robusta* embryos was removed by treatment with 0.1% trypsin (Thermo Fisher) in HL solution [S12]. *C. intestinalis* embryos were dechorionated with a solution of 1% sodium thioglycolate, 50 mM NaOH, and 2.5% Pronase [S2]. *C. cf. marginatus* embryos lack a chorion; eggs and embryos possess only a loose jelly, which was dispersed by aspiration or during fixation. Eggshells of *C. elegans* embryos were removed by freeze crack method [S10].

All eggs and embryos were fixed in either 100% methanol or 'von Dassow Fixative' (100 mM HEPES, 50mM EGTA, 10 mM $MgSO_4$, 500 mM dextrose, 2% formaldehyde, 0.2% glutaraldehyde, and 0.2% triton-X-100) [S13]. Embryos fixed in 100% methanol were rehydrated through a series of 75%, 50%, and 25% methanol solutions in 1X PBST. Embryos fixed in the von Dassow fixative were quenched with 0.1% sodium borohydride in 1X PBS. *X. laevis* embryos were bleached in a 10% H_2O_2 solution in methanol prior to immunostaining following previously described methods [S14]. After fixation, similar methods for immunostaining were used for all embryos. Briefly, embryos were washed in 1X PBST and blocked overnight at 4°C with 5% goat serum + 0.1% BSA in 1X PBST. Embryos were stained at 4°C with a monoclonal anti-α-tubulin antibody (Sigma, T9026) or YL1/2 monoclonal antibody (Immunologicals Direct) diluted to 1:1000 or 1:500 in 1X PBST. Embryos were washed with multiple changes of 1X PBST and then incubated at 4°C with goat anti-mouse Alexa Fluor 568 (Life Technologies) secondary antibody diluted 1:1000 in 1X PBST and Hoechst. Embryos were washed again with multiple changes of 1X PBST prior to mounting. This standard protocol was followed for all embryos, however blocking, antibody incubation, and

washing times differed: larger embryos had longer incubation and washing times, typically 24-72 hours, and smaller embryos had shorter incubation and washing times, typically 2-12 hours. Smaller embryos were mounted on a glass slide in vectashield (Vector Laboratories) or 50% glycerol and sealed with a coverslip and clear nail polish. Larger embryos (*X. laevis* and *H. robusta*) were cleared in 2:1 benzyl benzoate to benzyl alcohol and mounted in clearing solution on a glass slide surrounded by a vacuum grease cushion and sealed loosely with a coverslip. Imaging of embryos was carried out on one of three different microscopes: an Olympus BX51 fluorescence microscope with a 40x objective (Olympus UPlanFl N, Olympus; NA 0.75)**;** a Zeiss LSM 780 NLO AxioExaminer laser scanning confocal microscope equipped with 20x (Zeiss 1.0 W Plan Apo; NA 1.0) and 40x (Zeiss 1.4 Plan Apo; NA 1.4) objectives; a Olympus FluoView laser scanning confocal microscope equipped with a 60X objective (Plan Apo; NA 1.4). Type of the microscope used is indicated in final dataset (metadata).

Spindle Quantification

All measurements were done by manually tracing the spindle with a line (to measure lengths and widths) or a circle (to measure diameter), using either tubulin or DNA fluorescence, to quantify different parameters in ImageJ. The following parameters were measured for every spindle imaged: cell diameter (the largest diameter of the cell parallel to the pole-to-pole axis), aster-to-aster spindle length (the longest distance spanning the centers of the two centrosome asters), pole-to-pole spindle length (the longest distance between spindle poles defined as where dense interpolar microtubule arrays end), spindle width (the longest width of the interpolar microtubule array),

centrosome aster-to-pole distance (the longest distance between the centrosome aster and spindle pole), centrosome aster diameter (the largest diameter of the centrosome aster defined by circular pattern of bright tubulin fluorescence), astral microtubule array diameter (the largest diameter of the astral microtubule array as defined by the outer edge where most astral microtubules end), metaphase plate length (the longest length of the metaphase DNA mass defined as the vertical axis of the DNA perpendicular to the pole-to-pole axis), and metaphase plate width (the longest width of the metaphase DNA mass defined as the horizontal axis of the DNA parallel to the pole-to-pole axis). If a parameter/measurement was not clearly identifiable in an image, then no measurement was made. For all analysis, measurements were only included from spindles that were clearly in metaphase, as judged by DNA staining that showed clear congression of chromosomes to a metaphase plate. It should be noted that this study would not be possible without fixation and associated post-processing, which leaves the possibility of a size estimate artifact. Some of the measured cell diameters exceed known live dimensions for *S. purpuratus* and *A. miniata* by ~10%, which could be due to compression in mounting, fixation-associated swelling, or variability in egg batches. Given the diversity of materials, we did not attempt to impose any post-measurement screening or correction factors, and report measurements as originally assessed.

Quantification of microtubule intensity was performed using ImageJ as previously described [S15]. Briefly, a 30-pixel-wide linear region was manually overlaid along the center of pole-to-pole axis and the average tubulin fluorescence pixel intensity was calculated along the horizontal line. To capture a full profile of microtubule intensity along the spindle length, lines were drawn starting just outside the outermost edge of

one astral microtubule array, where the majority of astral microtubules terminated, and ending at the outermost edge of the other astral microtubule array. Intensity profiles were split into 100 bins and tubulin intensity was averaged for every bin along the length of the linescan. Meiotic and mitotic spindle intensity profiles from each species were averaged in this manner [S15].

Datasets and Data Analysis

All measurements are available as SQLite database and csv file, all data and code is available at http://emmaggie.github.io/SPINDLE_ZOO. There are 2001 measurements (rows) for 20 organisms, of 33 variables (columns), out of which 19 are continuous, 7 - categorical, 2 represent count data, 2 – time data, and 4 can be considered experimental metadata. Data was collected across 20 organisms. Dataset is not orthogonal. Some data entries are missing (~8%) and not all categories (e.g. organisms) are equally represented. All data was analyzed in Excel, R and Python. Throughout the analysis we used the correlation coefficient, r, to evaluate whether two variables correlate with each other. R^2 was used when fitting a statistical model to data, to get an estimate of the fraction of variation in response variable explained by the model.

Identification of linear scaling regime

To identify the cell size range within which linear spindle scaling occurred, we fitted a linear model to spindle size data over progressively smaller cell size ranges (in 10 μm steps). We monitored R^2 during the procedure, looking for the cell size range, which corresponds to the maximum fraction of spindle size variance explained (i.e. maximum

R^2). Across all mitotic spindles at metaphase, the cell size range corresponding to highest R^2 was up to ~110 μm for spindle lengths measured from pole-to-pole and 120 μm for aster-to-aster measurements (**Figure S2A**). To determine the cell size range within which linear spindle scaling occurred for each organism, this process was repeated for mitotic metaphase spindles of each individual species (**Figure 2B)**.

Piecewise regression

Given that our search for linear scaling regime would result in a number of data points not accounted for (i.e. all points associated with cell sizes above ~110-120 μm), we sought to apply a two component linear model to take advantage of all data points in an approach known as piecewise regression. To identify an optimal break point between the two components of the model, we applied the model across a range of cell diameter splits and monitored model performance using information criterion (Akaike and Bayesian). Cell diameters corresponding to the minimum in information criterion was chosen as a break point for the final model. When applied across all organisms, the two-component linear model accounts for 81% of total variation in pole-to-pole spindle length (90% in aster-to-aster) and identifies similar end points of linear scaling regime (110-120 μm, **Figure S2B**).

Tree models

To determine, which continuous variables should be included in a multiple regression model of meiotic/mitotic spindle length, we ran regression tree algorithms using all continuous variables in our data set [S16, S17]. Regression trees split a continuous response variable (here: spindle length) into subsets based on its relationship to explanatory variables (predictors) included in the model. In the

implementation we used, each split was based on the explanatory variable and its threshold, which leads to the greatest change in deviance explained (i.e. results in lowest unexplained variation at any given split). This was done through a search of all possible thresholds for all explanatory variables. After the split, the search was repeated in a similar manner within each branch. Splitting was ended when another split did not add much in terms of deviance explained (here: 1% of the null deviance) or when there was a low number of samples left at a particular branch (here: 6 samples). Therefore, regression trees allowed for identification of the strongest predictors of the response variable in multivariate data [S17]. The regression tree model accounted for approximately 90% of variation in mitotic spindle length and 79% in meiotic spindle length (90% if spindle aspect ratio is included in the model). Interestingly, in both cases (mitotic and meiotic), cell diameter alone accounted for the largest fraction of deviance explained: 55% in the case of mitotic spindles and 40% in the case of meiotic spindles. It is worth noting that the dataset for meiotic spindle is smaller and sparser, thus estimates inferred from it are likely less accurate. We used the variables selected by the regression tree to propose a multiple regression model for the continuous variables. The maximal model can be described with a function*:* $y=\beta_0+ \beta_1x_1+\ldots+ \beta_ix_{15}$, where y is meiotic/mitotic metaphase spindle length (pole to pole), x is the feature measured and β is the parameter estimated from the model. Variables were chosen according to the rank provided by the tree models and potential curvature in the behavior of variables was evaluated through generalized additive model for each variable. We used backward selection approach to eliminate non-significant terms from the model. The final multiple regression model for mitotic spindles accounted for 85% of variation in the spindle size

and utilized terms: cell diameter, inner aster diameter, metaphase plate length, square of genome size and square of inner aster diameter. In the case of meiotic spindles, the final model accounted for 83% of variation and included the terms: cell diameter, genome size, metaphase plate length, inner aster diameter, square of cell diameter, interaction terms: between cell diameter and spindle aspect ratio, genome size and spindle aspect ratio, spindle aspect ratio and metaphase plate length, cell diameter genome size and spindle aspect ratio. The multiple regression model for the meiotic size was less robust and strongly depended on the order of components removal in backward selection approach.

Unsupervised learning (PCA and clustering)

In a preprocessing step for unsupervised learning each feature (x) was scaled and centered (x_i*-mean(x))/sd(x)* and missing values filled through bagged trees method (lower rank matrix approximation was also tested). Scaled and centered data were subject to Principle Component Analysis and 7 components (accounting for ~91% of variance) selected for downstream analysis. Given the intrinsic caveats of our dataset (non-orthogonality, unequal representation of categorical variables), which might contribute to unequal cluster sizes, we decided to use affinity propagation [S18] to explore additional structure in our data. Even though the method converged to a reasonable number of clusters (4-6), they did not partition the data set in any way that we could interpret, given individual categorical variables we considered. We did not attempt to interpret the clusters in the context of combinations of categorical variables, and it is possible that clusters could be interpreted in such approach.

Supplemental References

S1. Sive, H.L., Grainger, R.M., and Harland, R.M. (2000). Early Development of Xenopus laevis: A Laboratory Manual (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

S2. Christiaen, L., Wagner, E., Shi, W., and Levine, M. (2009). Isolation of sea squirt (Ciona) gametes, fertilization, dechorionation, and development. Cold Spring Harb Protoc. *12,* pdb.prot5344.

S3. Gould, M.C. (1967). Echiuroid worms: Urechis. In Methods in Developmental Biology, F.H. Wilt and N.K. Wessells eds., (New York: Crowell) pp. 277-311.

S4. Gould, M.C., and Stephano, J.L. (1999). MAP kinase, meiosis, and sperm centrosome suppression in Urechis caupo. Dev. Biol. *1,* 348-358.

S5. Wessel, G.M., Reich, A.M., and Klatsky, P.C. (2010). Use of sea stars to study basic reproductive processes. Syst. Biol. Reprod. Med. *3,* 236-245.

S6. Leahy, P.S. (1986). Laboratory culture of Strongylocentrotus purpuratus Adults, Embryos, and Larvae. In Methods in Cell Biology: Echinoderm gametes and embryos, T.E. Schroeder ed., (Massachusetts: Academic Press) pp. 1-13.

S7. Sano, N., Obata, M., and Komaru, A. (2013). Mitochondrial DNA transcription levels during spermatogenesis and early development in doubly uniparental inheritance of the mitochondrial DNA system of the blue mussel Mytilus galloprovincialis. Zoolog Sci. *8,* 675-679.

S8. Costello, D.P. and Henley, C. (1971). Methods for obtaining and handling marine eggs and embryos (Woods Hole, MA: Marine Biological Laboratory).

S9. Maslakova, S.A. (2010). Development to metamorphosis of the nemertean pilidium larva. Front. Zool. *1,* 30-9994-7-30.

S10. Miller, D.M., and Shakes, D.C. (1995). Immunofluorescence microscopy. Methods Cell Biol. 365-394.

S11. Weisblat, D.A., and Kuo, D.H. (2009). Handling of Helobdella (leech) embryos. Cold Spring Harb Protoc. *4,* pdb.prot5189.

S12. Weisblat, D.A., and Kuo, D.H. (2009). Devitellinization of living Helobdella (leech) embryos. Cold Spring Harb Protoc. *4,* pdb.prot5191.

S13. Foe, V.E., and von Dassow, G. (2008). Stable and dynamic microtubules coordinately shape the myosin activation zone during cytokinetic furrow formation. J. Cell Biol. *3,* 457-470.

S14. Wuhr, M., Chen, Y., Dumont, S., Groen, A.C., Needleman, D.J., Salic, A., and Mitchison, T.J. (2008). Evidence for an upper limit to mitotic spindle length. Curr. Biol. *16,* 1256-1261.

S15. Wilbur, J.D., and Heald, R. (2013). Mitotic spindle scaling during Xenopus development by kif2a and importin alpha. Elife e00290.

S16. Kuhn, M., and Johnson, K. (2013). Applied predictive modeling Springer).

S17. Crawley, M. (2012). The R book Wiley).

S18. Bodenhofer, U., Kothmeier, A., and Hochreiter, S. (2011). APCluster: an R package for affinity propagation clustering. Bioinformatics *17,* 2463-2464.