Developmental Cell, Volume *37*

Supplemental Information

Co-evolving CENP-A and CAL1 Domains Mediate Centromeric CENP-A Deposition across *Drosophila* Species

Leah Rosin and Barbara G. Mellone

Supplemental Information

Supplemental Figures and Table Legends

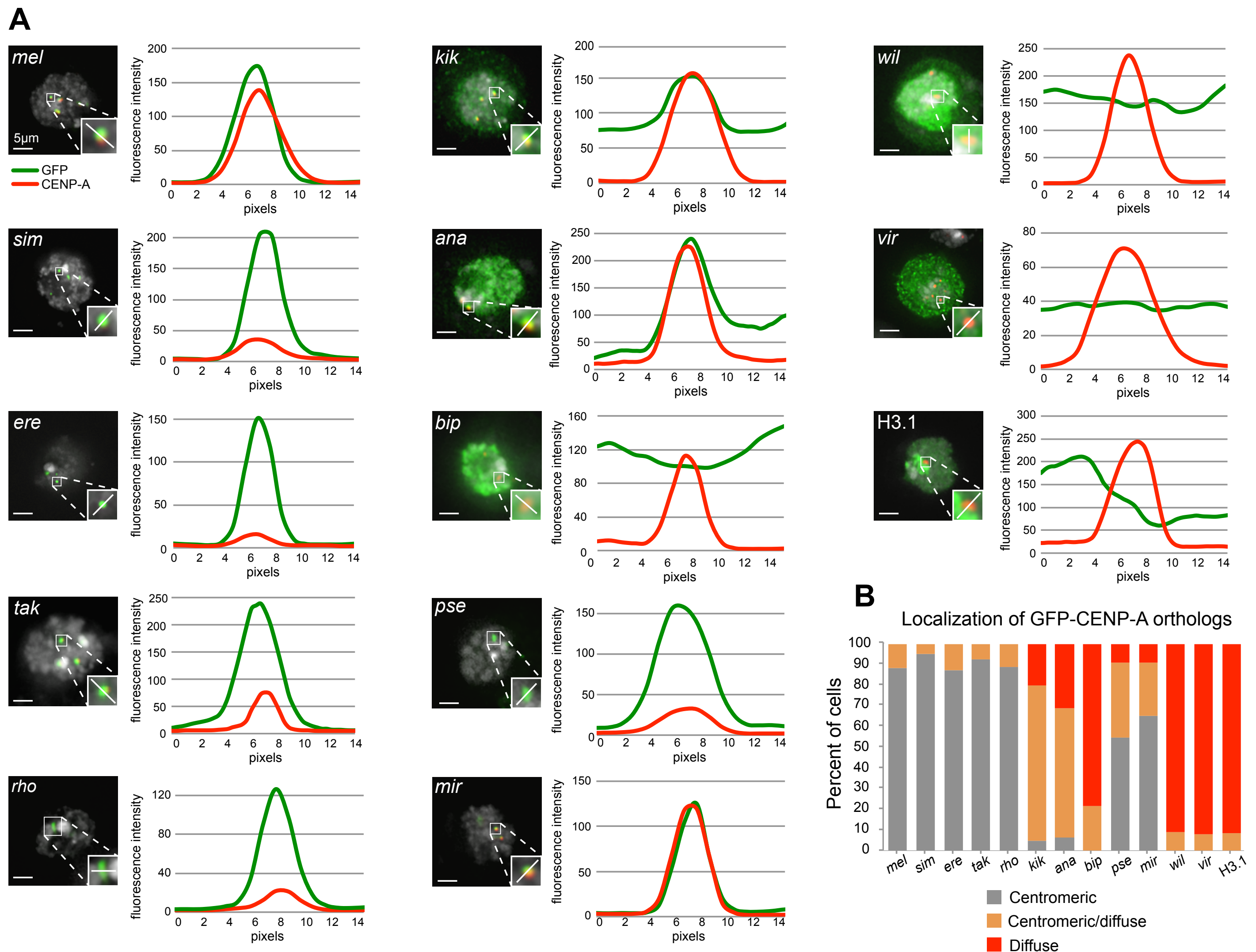


Figure S1, related to Figure 1. Line plots of GFP-tagged CENP-A orthologs and D. *melanogaster* CENP-A localization.
A) Representative IF images of interphase *mel* S2 cells transiently expressing GFP-tagged CENP-A orthologs from the indicated species and corresponding line plots showing the relative CENP-A (red) and GFP (green) fluorescence intensities. DAPI is shown in gray. GFP-H3.1 is shown as a control.
B) Quantification of line plots in A. Plots with clear GFP peaks overlapping with *mel* CENP-A peaks and low non-centromeric GFP signal (as in the examples for *mel, sim, ere, tak, rho, pse*) were scored as centromeric (gray bars). Plots displaying both GFP peaks as well as non-centromeric GFP signal (see *kik* and *ana*) were scored as centromeric/diffuse (orange bars), while plots showing no clear GFP peak overlapping with *mel* CENP-A (see as *bip, wil, vir, mel* H3.1) were scored as diffuse (red bars).

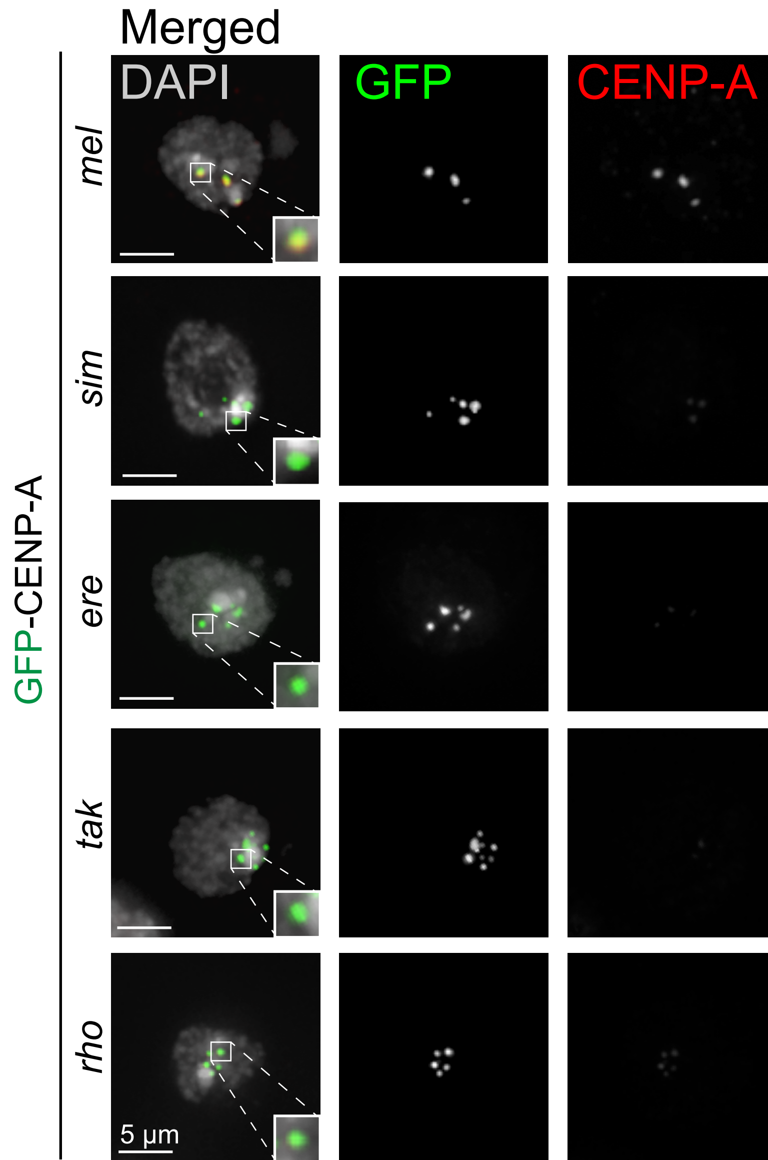


Figure S2, related to Figure 1. Centromeric localization of CENP-A orthologs results in decreased levels of *D. melanogaster* CENP-A in a subset of S2 cells.
IF images of interphase *mel* S2 cells transiently expressing GFP-tagged CENP-A orthologs from *mel*, *sim*, *ere*, *tak*, and *rho* in which endogenous CENP-A levels are noticeably low. Note that the *mel* CENP-A antibody (red) does not recognize CENP-A orthologs from *sim*, *ere*, *tak*, and *rho*. DAPI is shown in grey, GFP in green. Insets show magnified individual centromeres with merged colors.

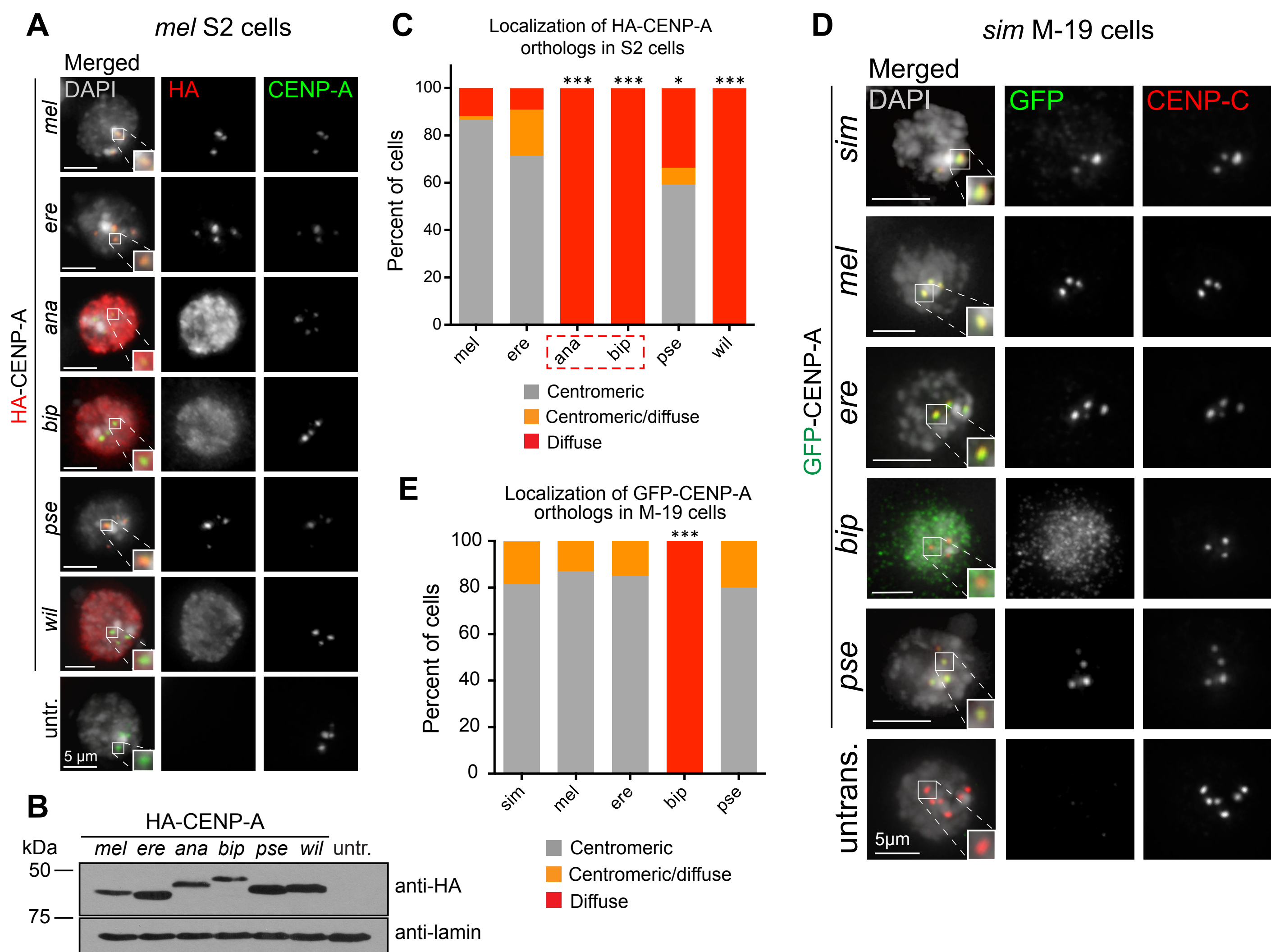


Figure S3, related to Figure 1. ***D. bipectinata*** **CENP-A does not localize to the centromeres of** ***D. melanogaster*** **or** ***D. simulans*** **cells.**

A) Representative IF images of interphase *mel* S2 cells transiently expressing HA-tagged CENP-A orthologs from *mel*, *ere*, *ana*, *bip*, *pse*, and *wil*. DAPI is shown in grey, HA in red, and *mel* CENP-A in green. Zoomed panels show individual centromeres with merged colors.

B) Western blots of total cell extracts showing the expression levels and protein sizes of HA-tagged CENP-A orthologs analyzed in A and B. Lamin antibody (loading control).

C) Quantification of the images shown in A. Images were manually classified as having either centromeric HA signal only (grey bars), diffuse HA signal (red bars), or centromeric and diffuse HA signal (orange bars). n = 50 transfected cells per condition on average. ***p<0.0001 for *ana* or *bip* CENP-A compared to *mel* CENP-A centromeric localization (Fisher's two-tailed test); *p=0.002 for *pse* CENP-A compared to *mel* CENP-A localization (Fisher's two-tailed test). These data were confirmed by one biological replicate with the HA-tag (data not shown) and two with the GFP-tagged constructs (Figure 1).

D) Representative IF images of *sim* M-19 cells transiently expressing GFP-tagged CENP-A orthologs from *mel, sim, ere, bip,* and *pse*. DAPI is shown in grey, GFP in green, and *sim* CENP-C in red. As our CENP-A antibody does not recognize *sim* CENP-A, the CENP-C antibody is used here to mark the native centromere locus. Zoomed panels show individual centromeres with merged colors.

E) Quantification of the images shown in D. Cells were manually classified as having either only centromeric GFP signal (grey bars), diffuse GFP signal (red bars), or centromeric and diffuse (orange

bars). n≥20 cells per condition. These results were confirmed by two biological replicates (data not shown). ***p<0.0001 for *bip* CENP-A compared to *sim* CENP-A centromeric localization (Fisher's two-tailed test).

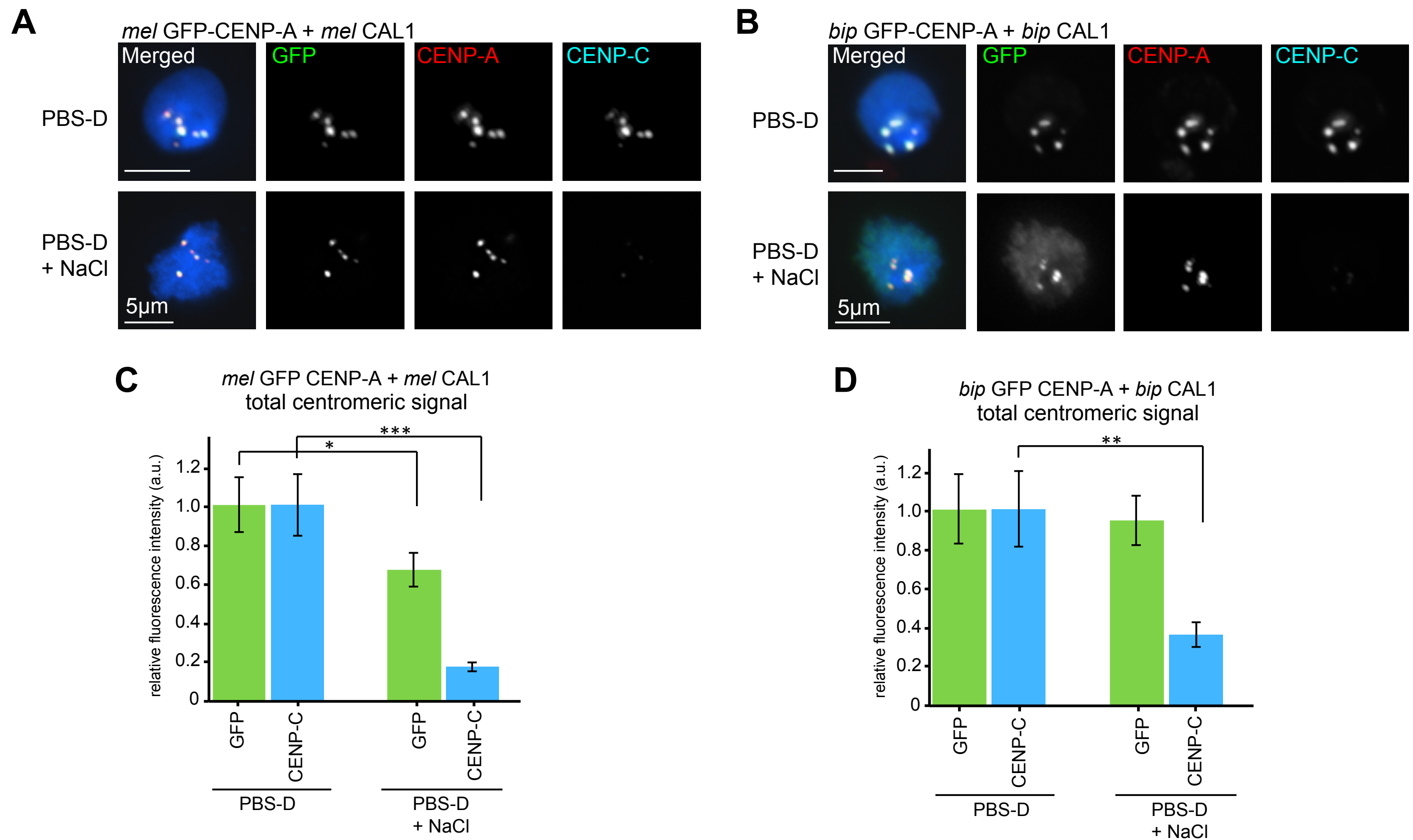


Figure S4, related to Figure 3. Centromere-localized *D. bipectinata* CENP-A is incorporated into chromatin.
A-B) Representative IF images of interphase *mel* S2 cells transiently co-expressing *mel* (A) or *bip* (B) GFP-CENP-A orthologs (green) and *mel* (A) or *bip* (B) HA-CAL1 (not shown). Cells were incubated with PBS-D (0.1% digitonin) plus or minus NaCl prior to fixation. CENP-C is expected to be extracted in the presence of NaCl and is used as a control (Perpelescu et al., 2009). The persistence of signal for GFP-CENP-A orthologs is indicative of chromatin incorporation. DAPI is shown in blue, *mel* CENP-A in red, and *mel* CENP-C in aqua.
C and D) Quantification of IF in A (C) and B (D). The relative fluorescence intensities of GFP CENP-A and CENP-C with and without NaCl incubation are shown. The error bars represent the standard error for an average of n=50 cells per condition. ***p<0.0001, **p=0.002, *p=0.019; unpaired t-test. This experiment was repeated with one biological replicate using PBS-T (0.1% triton) plus or minus NaCl (data not shown).

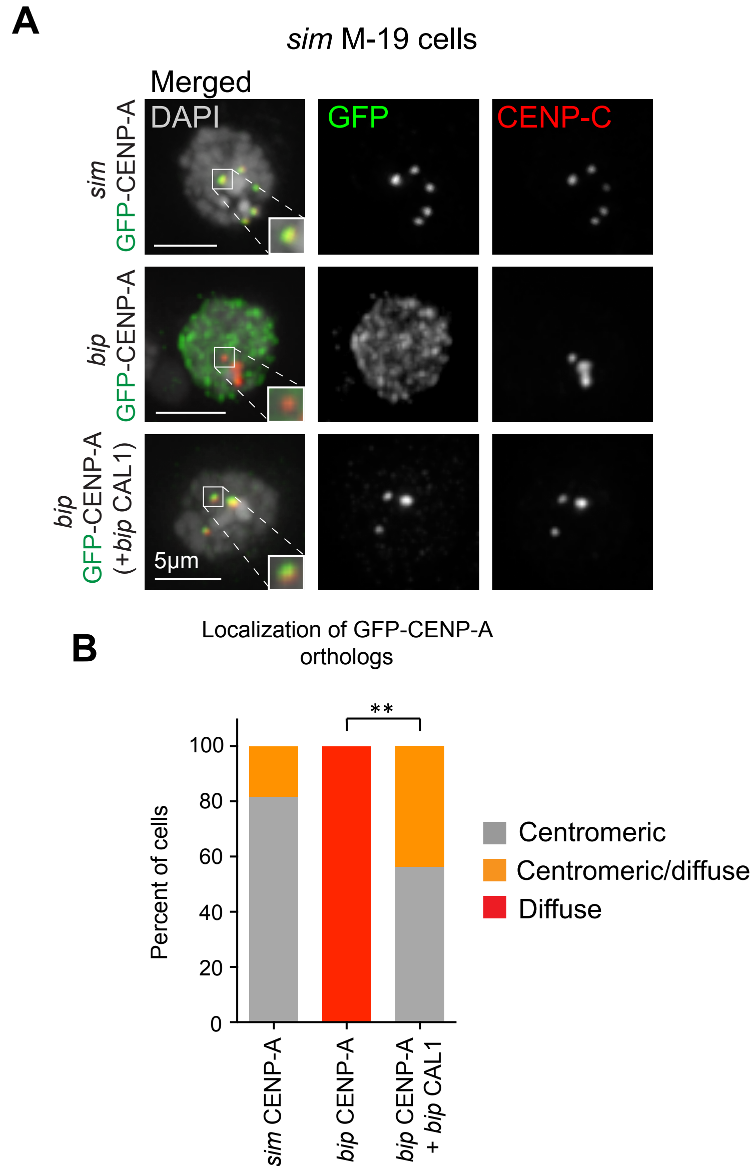
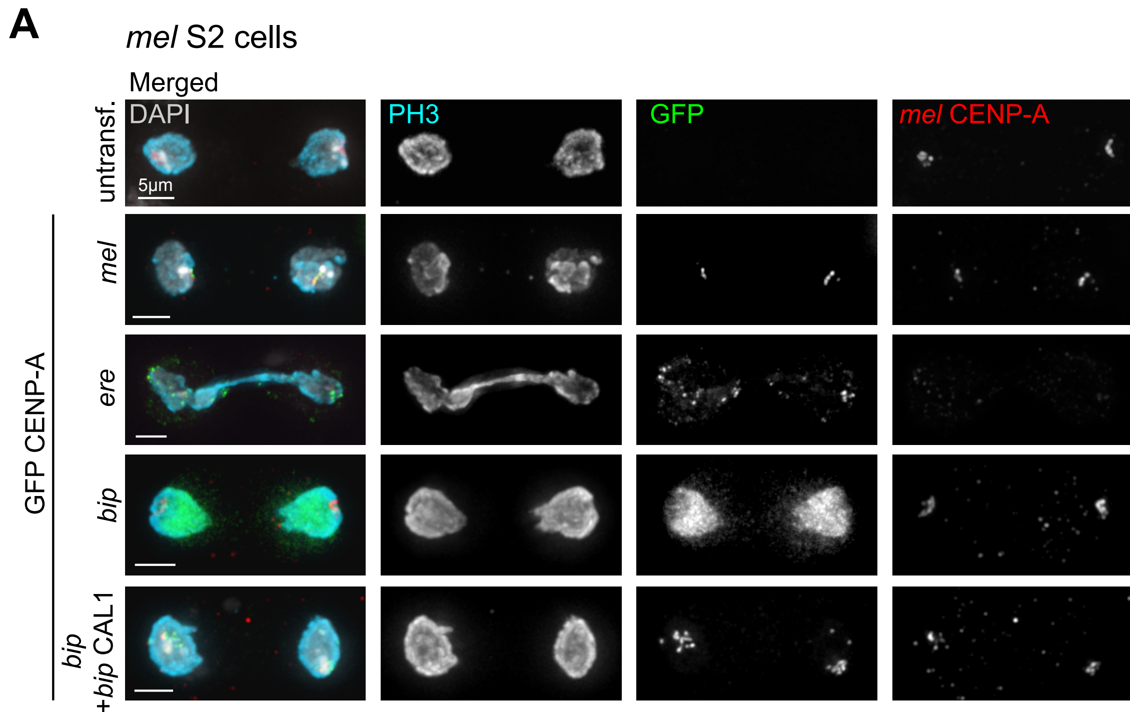


Figure S5, related to Figure 3. *D. bipectinata* CENP-A can localize to *D. simulans* centromeres when co-expressed with *D. bipectinata* CAL1.
A) Representative IF images of interphase *sim* M-19 cells transiently expressing *sim* or *bip* GFP-CENP-A alone (first and second rows, respectively), or *bip* GFP-CENP-A and *bip* HA-CAL1 (third row). DAPI is shown in grey, GFP in green, and CENP-C in red. Zoomed panels show individual centromeres with merged colors.
B) Quantification of the images shown in A. Cells were manually classified as having either exclusively centromeric GFP signal (grey bars), diffuse GFP signal (red bars), or centromeric and diffuse GFP signal (orange bars). $n \geq 30$ cells per condition. These data were confirmed by one biological replicate. **$p=0.0002$ (Fisher's two-tailed test) for the centromeric localization of *bip* CENP-A with and without *bip* CAL1 (data not shown).



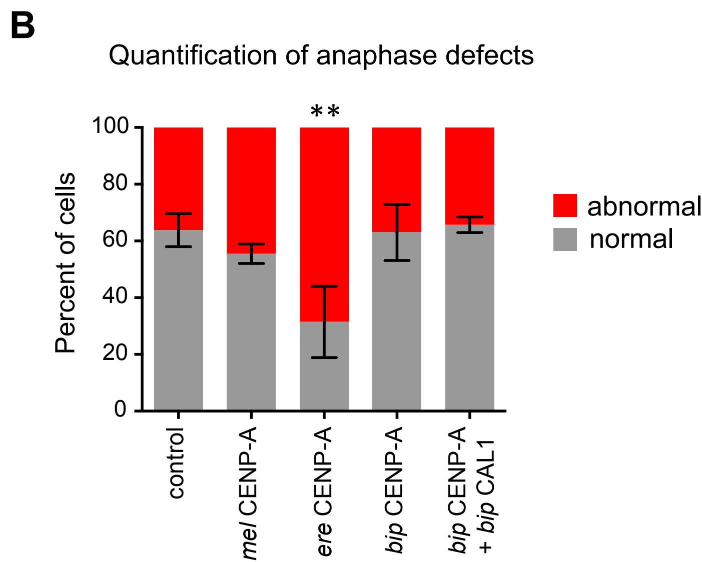


Figure S6, related to Figure 3. The *D. bipectinata* CENP-A/CAL1 complex is sufficient to nucleate *D. melanogaster* kinetochore formation

A) IF images of anaphase S2 cells transiently expressing GFP-CENP-A from *mel*, *ere*, or *bip* (second, third, and fourth column, respectively), or *bip* GFP-CENP-A and *bip* HA-CAL1 (fifth column). The first column shows untransfected S2 cells (GFP-negative). DAPI is shown in grey, GFP in green, *mel* CENP-A in red, and phosphorylated H3S10 (mitotic marker) in aqua.

B) Quantification of the images shown in B. Cells were manually classified as displaying normal (grey) or abnormal (stretched or lagging chromosomes; red) anaphases. The graph shows the average of three biological replicates, and the error bars are the standard deviation of the three biological replicates. n=131 cells total for untransfected images, n=157 for *mel* CENP-A, n=162 for *ere* CENP-A, n=76 for *bip* CENP-A, and n=90 for *bip* CENP-A with *bip* CAL1. **p=0.0009 (Fisher's two-tailed test of cells expressing *ere* CENP-A compared to cells expressing *mel* CENP-A.

Table S1, related to Experimental Procedures. Sources of genomic DNA used for PCR-cloning of *CENP-A* and *CAL1* orthologs.

Drosophila simulans	gDNA from Harmit Malik
Drosophila simulans	Tissue culture cells from DGRC, ML82-19a
Drosophila erecta	gDNA from Harmit Malik
Drosophila takahashii	Flies from UCSD Stock center, #14022-0311.13
Drosophila rhopaloa	Flies from UCSD Stock center, #14029-0021.01
Drosophila kikkawai	Flies from UCSD Stock center, #14028-0561.14
Drosophila ananassae	gDNA from Harmit Malik
Drosophila bipectinata	gDNA from Harmit Malik
Drosophila pseudoobscura	gDNA from Harmit Malik
Drosophila miranda	Flies from Doris Bachtrog
Drosophila wilistoni	Flies from Andy Clark
Drosophila virilis	Flies from Andy Clark

Table S2, related to Experimental Procedures. Primers used for cloning CENP-A and CAL1 orthologs, as well as the *mel* CENP-A^{bipL1} chimera.

mel_CENPA_ascI F	CAAAGGCGCGCCATGCCACGACACAGC
mel_CENPA_pacI R	CGGGTTAATTAACTAAAATTGCCGACCC
pse_CENPA_ascI_F	CAGTGGCGCGCCATGCGACCACCGACAAAAAACAG
pse_CENPA_pacI_R	CAGTTTAATTAATTAGTTAAAGCGACCATGGCTG
ere_CENPA_ascI_F	CAGTGGCGCGCCATGCCCCGACACAATGCTG
ere_CENPA_pacI_R	CAGTTTAATTAACTAGGCCAGCCGACCC
sim_CENPA_ascI_F	CAGTGGCGCGCCATGCCACGACACAGTAGAGCC
sim_CENPA_pacI_R	CAGTTTAATTAACTAAGCTTGCCGACCCCG
ana_CENPA_ascI_F	CAGTGGCGCGCCATGAGACCCCCACCAAAGC
ana_CENPA_pacI_R	CAGTTTAATTAATTAATTACGCCTCAAGTTGTCGC
vir_CENPA_ascI_F	CAGTGGCGCGCCATGCGTCCACGCACTG
vir_CENPA_pacI_R	CAGTTTAATTAATCAAAGATTACCATAGGTTTTGC
bip_CENPA_ascI_F	CAGTGGCGCGCCATGCGACCCCCACCAAAG
bip_CENPA_pacI_R	CAGTTTAATTAACTAGTTCGTCGCAAGGTTCTC
wil_CENPA_ascI_F	CAGTGGCGCGCCATGAGACCCCCTAGAGG
wil_CENPA_pacI_R	CAGTTTAATTAATCAATAGGCACTATCTTTGC
kik_CENPA_ascI_F	GAAAGGCGCGCCATGCGACCACCG
kik_CENPA_pacI_R	CGGGTTAATTAACTAGAGAAGACGATTATGAC
tak_CENPA_ascI_F	GAAAGGCGCGCCATGCCGAGAAAAAGTG
tak_CENPA_pacI_R	CGGGTTAATTAACTAGTTGTGACCCCGG
rho_CENPA_ascI_F	GAAAGGCGCGCCATGCCGAGGCAAG
rho_CENPA_pacI_R	CGGGTTAATTAACTAGAAATGACCCCGG
mir_CENPA_ascI_F	GAAAGGCGCGCCATGCGACCACCG
mir_CENPA_pacI_R	CGGGTTAATTAATTAGTTATATCGACAATGGC

ere_CAL1_ascI_F	CAGTGGCGCGCCATGGCGCAGGCGTT
ere_CAL1_pacI_R	CAGTTTAATTAATCAGTTGTCACCGGAATTATTC
bip_CAL1_ascI_F	CAGTGGCGCGCCATGTCTCAGGCACTGG
bip_CAL1_pacI_R	CAGTTTAATTAACTAGTTCTTCTCCAGAACAC
wil_CAL1_ascI_F	GAAAGGCGCGCCATGTCGTCGCACGTC
wil_CAL1_pacI_R	CGGGTTAATTAATTAGTTGTTTTTATTAGGCTT
pse_CAL1_ascI_F	CAGTGGCGCGCCATGTCGCATGCACTATTGG
pse_CAL1_pacI_R	CAGTTTAATTAATCAATCGTCTGTGGGATC
ere_CAL1_speI_F	CAGTACTAGTATGGCGCAGGCGTTGG
ere_CAL1_notI_R	ATTAGCGGCCGCGGTTGTCACCGGAATTATT
pse_CAL1_speI_F	CAGTACTAGTATGTCGCATGCACTATTGG
pse_CAL1_notI_R	ATTAGCGGCCGCTATCGTCTGTGGGATCC
bip_CAL1_speI_F	CAGTACTAGTATGTCTCAGGCACTGG
bip_CAL1_notI_R	ATTAGCGGCCGCGGTTCTTCTCCAGAACAC
bip_CAL1N_notI_R	CTTGGGCGGCCGCTACAGCCACTAGCTTG
bip_CAL1_XmaI_R	CAGTCCCGGGGTTCTTCTCCAGAACACTG
ana_CAL1_spe1_F	CAGTACTAGTATGTCGCAAGCACTGG
ana_CAL1_notI_R	ATTAGCGGCCGCGGTCCTTCTCCAGTACAC
wil_CAL1_speI_F	CGGGACTAGTATGTCGTCGCACGTCT
wil_CAL1_notI_R	ATTAGCGGCCGCaGTTGTTTTTATTAGGCTTCTCC
wil_CAL1_XmaI_R	CAGTCCCGGGGGTTGTTTTTATTAGGCTTCTC
CENPAmelL1Bip_F2n	CGTCTAGTGCGCGAGCTGCTTTACTCGCAAG
CENPAmelL1Bip_R1n	CTTGCGAGTAAAGCAGCTCGCGCACTAGACG
CENPAmelL1Bip_F3C	GTTCAAAATCTCCACCGGCGCCCTATTGGCC
CENPAmelL1Bip_R2c	GGCCAATAGGGCGCCGGTGGAGATTTTGAAC

Supplemental Experimental Procedures

Plasmids and Cloning

GFP-CENP-A constructs were generated by replacing *mel* CENP-A from the pCopia-LAP-CENP-A vector (Erhardt et al., 2008) with the PCR amplified CENP-A orthologs using AscI and PacI sites (New England Biolabs, Inc.). pCopia-flag-HA constructs were generated in the same way, using the pCopia-HA-CENP-A plasmid (Chen et al., 2014), and replacing *mel* CENP-A with either CENP-A or CAL1 orthologs using AscI and PacI sites. pMT-CAL1-GFP-LacI constructs were generated by replacing CENP-A from the pMT-CENP-A-GFP-LacI construct (Mendiburo et al., 2011) as previously described (Chen et al., 2014, 2015).

The *mel* CENP-A^{bipL1} chimera was created by sequential PCRs followed by cloning using the following primers (see Table S2 for sequences): mel_CENPA_AscI F, CENPAmelL1bip_F2n, CENPAmelL1bip_R1n, CENPAmelL1bip_F3C, CENPAmelL1bip_R2c, and mel_CENPA_PacI R. The *mel* CENP-A^{bipL1} chimera includes residues 1-159 of *mel* CENP-A, 219-232 of *bip* CENP-A, and 175-225 of *mel* CENP-A. Following digestion with AscI and PacI, it was cloned into the pCopia-LAP vector (Cheeseman and Desai, 2005; Erhardt et al., 2008).

The *bip* CENP-A^{melL1} chimera was synthesized as a gBlock (Integrated DNA Technologies), with flanking AscI and PacI sites for subcloning into the pCopia-HA vector.

The *bip*-*mel* CAL1 chimeras were synthesized as gBlocks with flanking SpeI and NotI sites (New England Biolabs, Inc.) for ligation into the pMT-GFP-LacI vector (Chen et al., 2014; Mendiburo et al., 2011). The *mel* CAL1$^{bip1\text{-}160}$ construct includes residues 1-173 of *bip* CAL1 and residues 163-407 of *mel* CAL1. The *mel* CAL1$^{bip1\text{-}40}$ construct includes residues 1-40 of *bip* CAL1 and residues 41-407 of *mel* CAL1. The *mel* CAL1$^{bip41\text{-}160}$ construct includes residues 1-40 of *mel* CAL1, residues 41-173 of *bip* CAL1, and residues 41-407 of *mel* CAL1. All constructs were verified by sequencing.

Quantification

For experiments in *mel* S2 cells, the localization of GFP or HA-tagged CENP-A or CAL1 orthologs was quantified manually and constructs were classified as either fully centromeric (GFP- or HA-signal co-localized with *mel* CENP-A), diffuse (exclusively euchromatic or cytoplasmic), or centromeric/diffuse (localizing to centromeres and throughout the chromatin). For experiments performed in *mel* lacO S2 cell lines, images were analyzed manually for the presence of the LacI fusion protein on the arms of chromosomes 2 or 3. Only chromosomes showing ectopic GFP (or ectopic CENP-A when using CAL1-LacI without GFP; data not shown) were scored for the presence or absence of centromere/kinetochore proteins. Transient co-transfections in *mel* lacO S2 cells resulted in low n-values (where n equals the number of lacO-positive co-transfected cells) per experiment, due to the polyclonal nature of this cell line combined with transfection efficiency, but 3 biological replicates were performed for each experiment, or 2 when the experiment was repeated with both HA- and GFP-tagged constructs.

Transfection efficiencies in *sim* M-19 cells were significantly lower than in S2 cells, resulting in low n-values (where n equals the number of GFP positive cells) per experiment. 3 biological replicates were performed for each experiment.

For salt extractions, the total centromeric GFP or CENP-C fluorescence signal was found using SoftWorx software (Applied Precision).

Western blots of IPs were quantified using Image Studio Software (Li-Cor) using the Shapes/Quantification function. Intensity of GFP-CENP-A in the IP sample was normalized relative to that of the respective input sample.

Line plots

IF images were quantitatively analyzed in ImageJ using the lines function to create line plots of the endogenous CENP-A and GFP CENP-A fluorescence intensity. Line plots of individual cells were then classified as having diffuse GFP-CENP-A fluorescence (broad signal with no distinct centromeric peak), centromeric/diffuse GFP-CENP-A fluorescence (broad signal with a peak overlapping endogenous *mel* CENP-A peak), or centromeric GFP-CENP-A localization (low background with a distinct centromeric peak).

Multiple sequence alignments

Multiple sequence alignments were performed using the Geneious ® 8.1.5 software (Kearse et al., 2012) using a BLOSUM80 (Eddy, 2004; Henikoff and Henikoff, 1992) cost matrix with free end gaps.

Statistical analyses

All statistical tests were performed using GraphPad ® scientific software. All p-values, with the exception of Figure S4, were calculated using a Fisher's two-tailed test. For Figure S4, p-values were calculated using an unpaired t-test (O'Mahony, 1986).

Supplemental References

Cheeseman, I.M., and Desai, A. (2005). A Combined Approach for the Localization and Tandem Affinity Purification of Protein Complexes from Metazoans. Sci. Signal. *2005*, pl1–pl1.

Chen, C.-C., Dechassa, M.L., Bettini, E., Ledoux, M.B., Belisario, C., Heun, P., Luger, K., and Mellone, B.G. (2014). CAL1 is the Drosophila CENP-A assembly factor. J. Cell Biol. *204*, 313–329.

Chen, C.-C., Bowers, S., Lipinszki, Z., Palladino, J., Trusiak, S., Bettini, E., Rosin, L., Przewloka, M.R., Glover, D.M., O'Neill, R.J., et al. (2015). Establishment of Centromeric Chromatin by the CENP-A Assembly Factor CAL1 Requires FACT-Mediated Transcription. Dev. Cell *34*, 73–84.

Eddy, S.R. (2004). Where did the BLOSUM62 alignment score matrix come from? Nat. Biotechnol. *22*, 1035–1036.

Erhardt, S., Mellone, B.G., Betts, C.M., Zhang, W., Karpen, G.H., and Straight, A.F. (2008). Genome-wide analysis reveals a cell cycle-dependent mechanism controlling centromere propagation. J. Cell Biol. *183*, 805–818.

Henikoff, S., and Henikoff, J.G. (1992). Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. U. S. A. *89*, 10915–10919.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinforma. Oxf. Engl. *28*, 1647–1649.

Mendiburo, M.J., Padeken, J., Fülöp, S., Schepers, A., and Heun, P. (2011). Drosophila CENH3 Is Sufficient for Centromere Formation. Science *334*, 686–690.

O'Mahony, M. (1986). Sensory evaluation of food: statistical methods and procedures (New York: M. Dekker).

Perpelescu, M., Nozaki, N., Obuse, C., Yang, H., and Yoda, K. (2009). Active establishment of centromeric CENP-A chromatin by RSF complex. J. Cell Biol. *185*, 397–407.