Published in final edited form as:
Science. 2013 June 7; 340(6137): 1211–1214. doi:10.1126/science.1234393.

Stepwise evolution of essential centromere function in a *Drosophila* neogene

Benjamin D. Ross[1,2], **Leah Rosin**[3,*], **Andreas W. Thomae**[4,*], **Mary Alice Hiatt**[2,5,*], **Danielle Vermaak**[2,6,*], **Aida Flor A. de la Cruz**[2,7], **Axel Imhof**[4], **Barbara G. Mellone**[3], and **Harmit S. Malik**[2,7,**]

[1]Molecular and Cellular Biology Program, University of Washington, Seattle, WA USA

[2]Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

[3]Department of Molecular and Cell Biology, University of Connecticut, Storrs CT

[4]Munich Centre of Integrated Protein Science and Adolf-Butenandt Institute, Ludwig Maximilians University of Munich, Munich, Germany

[7]Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

Abstract

Evolutionarily young genes that serve essential functions represent a paradox since they must perform a function that either was not required until after their birth or was redundant with another gene. How young genes rapidly acquire essential function is largely unknown. Here, we trace the evolutionary steps by which the *Drosophila* gene *Umbrea* acquired an essential role in chromosome segregation in *D. melanogaster*, since its origin less than 15 million years ago. Umbrea neofunctionalization occurred via loss of an ancestral heterochromatin-localizing domain, followed by alterations that rewired its protein interaction network and led to species-specific centromere localization. Our evolutionary cell biology approach provides temporal and mechanistic detail into how young genes gain essential function. Such innovations may constantly alter the repertoire of centromere proteins in eukaryotes.

Young essential genes (1) challenge long-standing dogmas about the relationship between essentiality and conservation (2). Partitioning of essential, ancestral functions (subfunctionalization) between (old) parental and (young) daughter genes (3, 4) explains one route by which young genes become essential. More difficult to understand is how new genes become essential via the emergence of novel function (neofunctionalization) (5). This could result from partial duplication of ancestral genes, novel gene fusions, or by rapid amino acid changes (6). The contribution of each of these processes to the acquisition of essential function is unknown, as are the underlying molecular changes.

**address correspondence to: Harmit S. Malik, 1100 Fairview Avenue N A2-025, Seattle, WA 98109. Tel: (206) 667-5204; Fax: (206) 667-6522; hsmalik@fhcrc.org.
*these authors contributed equally to this work
[5]Law firm of Fitzpatrick, Cella, Harper & Scinto, New York, NY 10104
[6]Present address: Science-wise Consulting, Seattle, WA 98109

To gain insight into the birth and evolution of essential function, we focused on one newly evolved gene in *Drosophila*. *Umbrea/HP6/CG15636* arose via duplication of the intronless *Heterochromatin Protein 1B* (*HP1B*) gene into an intron of the *dumpy* gene (Fig. 1A)(7). HP1B is a chromosomal protein, which predominantly localizes to heterochromatin in *D. melanogaster* cells and regulates gene expression(8). *HP1B* is dispensable for viability (8), yet *Umbrea* is essential in *D. melanogaster* (1, 9) based on RNAi knockdown phenotypes. The 100% late larval-pupal lethality upon *Umbrea* knockdown could be rescued by an *Umbrea-GFP* fusion (Fig. S1). Genetic knockout experiments (Fig. S1) further confirmed that *Umbrea* is essential in *D. melanogaster*.

We traced *Umbrea's* evolutionary path following duplication from *HP1B* to understand when and how essential function was gained by comparing the localization of HP1B and Umbrea in *D. melanogaster* Kc cells. GFP-tagged HP1B proteins from both *D. melanogaster* and *D. ananassae* (whose divergence predates birth of *Umbrea* (7)), localized to pericentric heterochromatin and euchromatin (Figs. 1B and S2). In contrast, Umbrea-GFP predominantly localized to interphase centromeres, but not telomeres (Fig. 1C, S3A–B). Specific antibodies raised against Umbrea (Fig S4A) confirmed its centromeric localization in developing spermatocytes and larval imaginal discs (Fig 1D–E and S4B–C).

On the basis of *Umbrea's* essentiality and centromere localization, we hypothesized that Umbrea was required for chromosome segregation. Upon depletion of of Umbrea by RNAi knockdown (Fig. S5A), *D. melanogaster* S2 cells displayed increased mitotic errors, including delayed chromosome alignment, early anaphase onset, lagging anaphase chromosomes, and multipolar configurations, compared to control cells ($p<.05$) (Fig. 1F–G, S5B, Movies S1-S3). These data suggest that Umbrea promotes proper chromosome segregation, but is not required for the localization of the centromeric histone Cid (Fig. 1F).

To date the birth of *Umbrea* and subsequent changes, we sequenced the *Umbrea* locus from 32 *Drosophila* species (Fig. S6A). While *HP1B* was preserved (7), we found *Umbrea* in only 20 of 32 species, dating its monophyletic origin to 12–15 million years ago (Fig. 2A, S6B). Using maximum likelihood methods, we observed evidence of both episodic and recurrent positive selection acting on *Umbrea* (Fig. S7A–D). These findings, together with the altered localization, lead us to conclude that neofunctionalization, not subfunctionalization, drove the divergence of *Umbrea* (10). Although *Umbrea* is essential in *D. melanogaster*, it was lost at least three independent times, in *D. fuyamai*, *D. eugracilis*, and in the *suzukii* clade (Fig. 2A), suggesting that *Umbrea* was not essential at or immediately following its birth.

Four lineages retained full-length *Umbrea* genes, two of which preserved an intact chromodomain (CD) and ancestral residues essential for binding histone H3 tri-methyl lysine 9 (H3K9me) (Fig. S8) (11). However, most extant *Umbreas* have lost their CDs, and retained only the chromoshadow domain (CSD), which mediates protein-protein interactions (12) (Fig. 2A). We first tested how CD loss affected HP1B function. We found that an HP1B-GFP fusion lacking the CD lost heterochromatin localization (Fig. 2B), consistent with the requirement of HP1-CD for H3K9me binding (13). Furthermore, fusion of the HP1B CD and hinge to Umbrea-GFP reverted localization from centromeres to

heterochromatin (Fig. 2C), suggesting that loss of the ancestral CD was necessary for Umbrea to gain new function. Our findings support a model of neofunctionalization that is facilitated via intermediate loss-of-function (14). Although CD loss was necessary, it was not sufficient for Umbrea neofunctionalization; both full-length (*D. fuyamai*) and CSD-only (*D. eugracilis* and the *suzukii* clade) *Umbrea* genes have been lost in evolution.

We next investigated the consequences of evolution in the Umbrea-CSD. CSDs are only found in HP1-family proteins and mediate interactions with other HP1s or proteins possessing degenerate PxVxL motifs (15). An amino acid alignment of HP1B and Umbrea revealed conservation of residues defining the CSD structural fold (Fig. 3A). In contrast, three of the nine residues that mediate specificity for PxVxL-recognition (16) changed along the branch leading to the *melanogaster* species subgroup (Fig. 3A, S9). We found that *D. melanogaster* Umbrea-CSD localized to centromeres (Fig. 3B). This property was not shared amongst HP1B- or even other Umbrea-CSDs, since neither 'parental' HP1B^{mel}-CSD nor Umbrea^{ptak}-CSD could localize to centromeric regions in *D. melanogaster* cells (Fig. S10B, 3C). We conclude that a discrete transition for centromere localization occurred in Umbrea-CSD after divergence of the *melanogaster* and *takahashii* subgroups, coincident with changes in the PxVxL-recognition residues. Indeed, reversion of these three residues (C15, I57, F59, Fig 3A, S9) to the ancestral state delocalized Umbrea^{mel}-CSD from centromeres (Fig. 3D). Moreover, replacement of the same residues in Umbrea^{ptak}-CSD to corresponding residues in Umbrea^{mel} resulted in a gain of centromere localization (Fig. 3E). These results suggest that centromeric localization by Umbrea-CSD originated in the common ancestor of the *melanogaster* species subgroup 5–7 million years ago. Consistent with this, we find that GFP-Umbrea^{tei} localizes to centromeres in *D. teissieri* cells (Fig. 3F). Centromeric localization may have also coincided with gain of essentiality, as Umbrea was lost three times prior to, but not after, CSD modification (Fig. 2A).

To test the prediction that mutation of PxVxL recognition resulted in CSD centromere localization by alteration of protein interactions, we performed proteomic analyses to identify proteins that co-immunoprecipitate with Umbrea in S2 cells (Fig. 3G). Many chromatin factors were found in this set (Table S1), including heterochromatin proteins HP4/Hip and HP5 (previously shown to be direct interactors of Umbrea(9, 17)), as well as novel interactions with the H3K9 methyltransferase Su(var)3–9 and the centromeric protein Cenp-C. Importantly, we found no overlap with protein partners of HP1B, which include the euchromatic proteins HP1C, Woc and Row (18) (Fig. 3H), suggesting a rewiring of the protein interaction network of Umbrea.

Our evolutionary analyses (Fig. S7A–D) indicated that the most recent innovations in Umbrea occurred in the short tail sequences that flank the CSD. We tested how these changes contributed to Umbrea neofunctionalization. While HP1B^{mel}-CSD alone showed no discrete localization (Fig. 2B), addition of Umbrea^{mel}-tails was sufficient to confer centromere localization (Fig. 4A). These data indicate that Umbrea may target centromeres using both the CSD and the tails. While the CSD likely mediates its localization via protein-protein interactions, Umbrea-tails may bind centromeric nucleic acids, analogous to the hinge region of mammalian HP1alpha, which binds DNA *in vitro* (19). Since centromeric DNA sequence diverges rapidly (20), we tested whether rapid evolution of the Umbrea-tails

resulted in species-specificity. We found that Umbrea^{sim} localized (Fig. 4B) to centromeres in *D. melanogaster*. However, Umbrea^{tei} and Umbrea^{yak} did not (Fig. 4C–D), localizing instead to distinct foci. Although positive selection of *Umbrea* preceded its centromeric localization (Fig. S7), these data suggest that positive selection in the *melanogaster* species subgroup resulted in species-specific centromere targeting, reminiscent of CenH3/Cid in *Drosophila* (21). For example, despite mislocalizing in *D. melanogaster* cells, Umbrea^{tei} appropriately localized to *D. teissieri* centromeres (Fig. 3F).

Our analyses suggest that gain of essential function evolved in discrete steps (Fig. 4E) (5) that involved the loss of an ancestral domain (CD), rewiring of protein interaction networks (CSD), and species-specific changes (tails). *Umbrea* was likely not essential for much of its evolutionary history, since intermediate forms were lost multiple times.

Our finding that *Umbrea* rapidly became essential for the conserved process of chromosome segregation is unexpected. *Drosophila* species that never possessed or lost *Umbrea* still carry out chromosome segregation. This suggests that the essential function of *Umbrea* might be a result of a lineage-specific requirement. Just as genetic conflicts arising during meiosis may drive rapid evolution of existing centromere proteins (22), we propose that recurrent changes at centromeric DNA satellites could drive the retention of duplicate genes like *Umbrea* to alleviate selective pressure on essential centromeric proteins. This is analogous to pathogen-driven genetic conflict, which promotes the diversification of existing and new antiviral immune genes (23). This process would result in idiosyncratic retention of centromeric proteins that become essential as they integrate into existing networks. Intriguingly, other HP1B-derived CSD-only genes are found in other *Drosophila* species that diverged prior to the birth of *Umbrea* (7), raising the possibility of convergent evolution of Umbrea-like centromere factors. This process may explain the broad diversity and divergence amongst centromere proteins across taxa (24). While a large fraction of the many young, essential genes identified in *Drosophila* (1) may result from subfunctionalization, others (like *Umbrea*) may illuminate other essential processes that could require recurrent genetic innovation to mitigate previously unappreciated adaptive challenges within the cell.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank J. Bloom, M. Daugherty, M. Emerman, D. Gottschling, M. Levine, M. Patel, N. Phadnis, K. Peichel and W. Shou for helpful comments and Drosophila colleagues for generous sharing of reagents. This work was supported by NIH Training grant T32HG000035 and an NSF GRFP (to BDR), an EU network grant (EpiGeneSys 257082, to AI), NSF award 1024973 (to BGM), a grant from the Mathers Foundation and grant R01GM074108 from the NIH (to HSM). HSM is an Early Career Scientist of HHMI. *Umbrea* DNA sequences have been submitted to GenBank under accession numbers KC660086- KC660100. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (PRIDE partner repository dataset identifier PXD000163).

References

1. Chen S, Zhang YE, Long M. New genes in Drosophila quickly become essential. Science. 2010; 330:1682. [PubMed: 21164016]
2. Miklos GL, Rubin GM. The role of the genome project in determining gene function: insights from model organisms. Cell. 1996; 86:521. [PubMed: 8752207]
3. Force A, et al. Preservation of duplicate genes by complementary, degenerative mutations. Genetics. 1999; 151:1531. [PubMed: 10101175]
4. Stoltzfus A. On the possibility of constructive neutral evolution. J Mol Evol. 1999; 49:169. [PubMed: 10441669]
5. Innan H, Kondrashov F. The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet. 2010; 11:97. [PubMed: 20051986]
6. Zhang J, Dean AM, Brunet F, Long M. Evolving protein functional diversity in new genes of Drosophila. Proc Natl Acad Sci U S A. 2004; 101:16246. [PubMed: 15534206]
7. Levine MT, et al. Phylogenomic Analysis Reveals Dynamic Evolutionary History of the Drosophila Heterochromatin Protein 1 (HP1) Gene Family. PLoS Genet. 2012; 8:e1002729. [PubMed: 22737079]
8. Zhang D, Wang D, Sun F. Drosophila melanogaster heterochromatin protein HP1b plays important roles in transcriptional activation and development. Chromosoma. 2011; 120:97. [PubMed: 20857302]
9. Joppich C, Scholz S, Korge G, Schwendemann A. Umbrea, a chromo shadow domain protein in Drosophila melanogaster heterochromatin, interacts with Hip, HP1 and HOAP. Chromosome Res. 2009; 17:19. [PubMed: 19190990]
10. Hahn MW. Distinguishing Among Evolutionary Models for the Maintenance of Gene Duplicates. J Hered. 2009; 100:605. [PubMed: 19596713]
11. Jacobs SA, Khorasanizadeh S. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. Science. 2002; 295:2080. [PubMed: 11859155]
12. Murzina N, Verreault A, Laue E, Stillman B. Heterochromatin dynamics in mouse cells: interaction between chromatin assembly factor 1 and HP1 proteins. Mol Cell. 1999; 4:529. [PubMed: 10549285]
13. Nakayama J, Rice JC, Strahl BD, Allis CD, Grewal SI. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. Science. 2001; 292:110. [PubMed: 11283354]
14. Ohta T. Evolution by gene duplication and compensatory advantageous mutations. Genetics. 1988; 120:841. [PubMed: 3224809]
15. Smothers JF, Henikoff S. The HP1 chromo shadow domain binds a consensus peptide pentamer. Curr Biol. 2000; 10:27. [PubMed: 10660299]
16. Thiru A, et al. Structural basis of HP1/PXVXL motif peptide interactions and HP1 localisation to heterochromatin. EMBO J. 2004; 23:489. [PubMed: 14765118]
17. Giot L, et al. A protein interaction map of Drosophila melanogaster. Science. 2003; 302:1727. [PubMed: 14605208]
18. Abel J, Eskeland R, Raffa GD, Kremmer E, Imhof A. Drosophila HP1c is regulated by an auto-regulatory feedback loop through its binding partner Woc. PLoS One. 2009; 4:e5089. [PubMed: 19352434]
19. Meehan RR, Kao CF, Pennings S. HP1 binding to native chromatin in vitro is determined by the hinge region and not by the chromodomain. EMBO J. 2003; 22:3164. [PubMed: 12805230]
20. Lohe AR, Brutlag DL. Identical satellite DNA sequences in sibling species of Drosophila. J Mol Biol. 1987; 194:161. [PubMed: 3112413]
21. Vermaak D, Hayden HS, Henikoff S. Centromere targeting element within the histone fold domain of Cid. Mol Cell Biol. 2002; 22:7553. [PubMed: 12370302]
22. Malik HS, Henikoff S. Major evolutionary transitions in centromere complexity. Cell. 2009; 138:1067. [PubMed: 19766562]
23. Daugherty MD, Malik HS. Rules of engagement: molecular insights from host-virus arms races. Annu Rev Genet. 2012; 46:677. [PubMed: 23145935]

24. Cheeseman IM, Desai A. Molecular architecture of the kinetochore-microtubule interface. Nat Rev Mol Cell Biol. 2008; 9:33. [PubMed: 18097444]
25. Prud'homme B, et al. Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. Nature. 2006; 440:1050. [PubMed: 16625197]
26. Henikoff S, Ahmad K, Malik HS. The centromere paradox: stable inheritance with rapidly evolving DNA. Science. 2001; 293:1098. [PubMed: 11498581]
27. Larkin MA, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007; 23:2947. [PubMed: 17846036]
28. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 2006; 34:W609. [PubMed: 16845082]
29. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007; 24:1586. [PubMed: 17483113]
30. Cox J, et al. Andromeda: a peptide search engine integrated into the MaxQuant environment. J Proteome Res. 2011; 10:1794. [PubMed: 21254760]
31. Voog J, D'Alterio C, Jones DL. Multipotent somatic stem cells contribute to the stem cell niche in the Drosophila testis. Nature. 2008; 454:1132. [PubMed: 18641633]
32. Wong B. Color blindness. Nat Methods. 2011; 8:441. [PubMed: 21774112]
33. Blower MD, Karpen GH. The role of Drosophila CID in kinetochore formation, cell-cycle progression and heterochromatin interactions. Nat Cell Biol. 2001; 3:730. [PubMed: 11483958]
34. Heeger S, et al. Genetic interactions of separase regulatory subunits reveal the diverged Drosophila Cenp-C homolog. Genes Dev. 2005; 19:2041. [PubMed: 16140985]
35. Klattenhoff C, et al. The Drosophila HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. Cell. 2009; 138:1137. [PubMed: 19732946]
36. Gao G, et al. HipHop interacts with HOAP and HP1 to protect Drosophila telomeres in a sequence-independent manner. EMBO J. 2010; 29:819. [PubMed: 20057353]
37. Heun P, et al. Mislocalization of the Drosophila centromere-specific histone CID promotes formation of functional ectopic kinetochores. Dev Cell. 2006; 10:303. [PubMed: 16516834]
38. Vizcaino JA, et al. The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. Nucleic Acids Res. 2013; 41:D1063. [PubMed: 23203882]

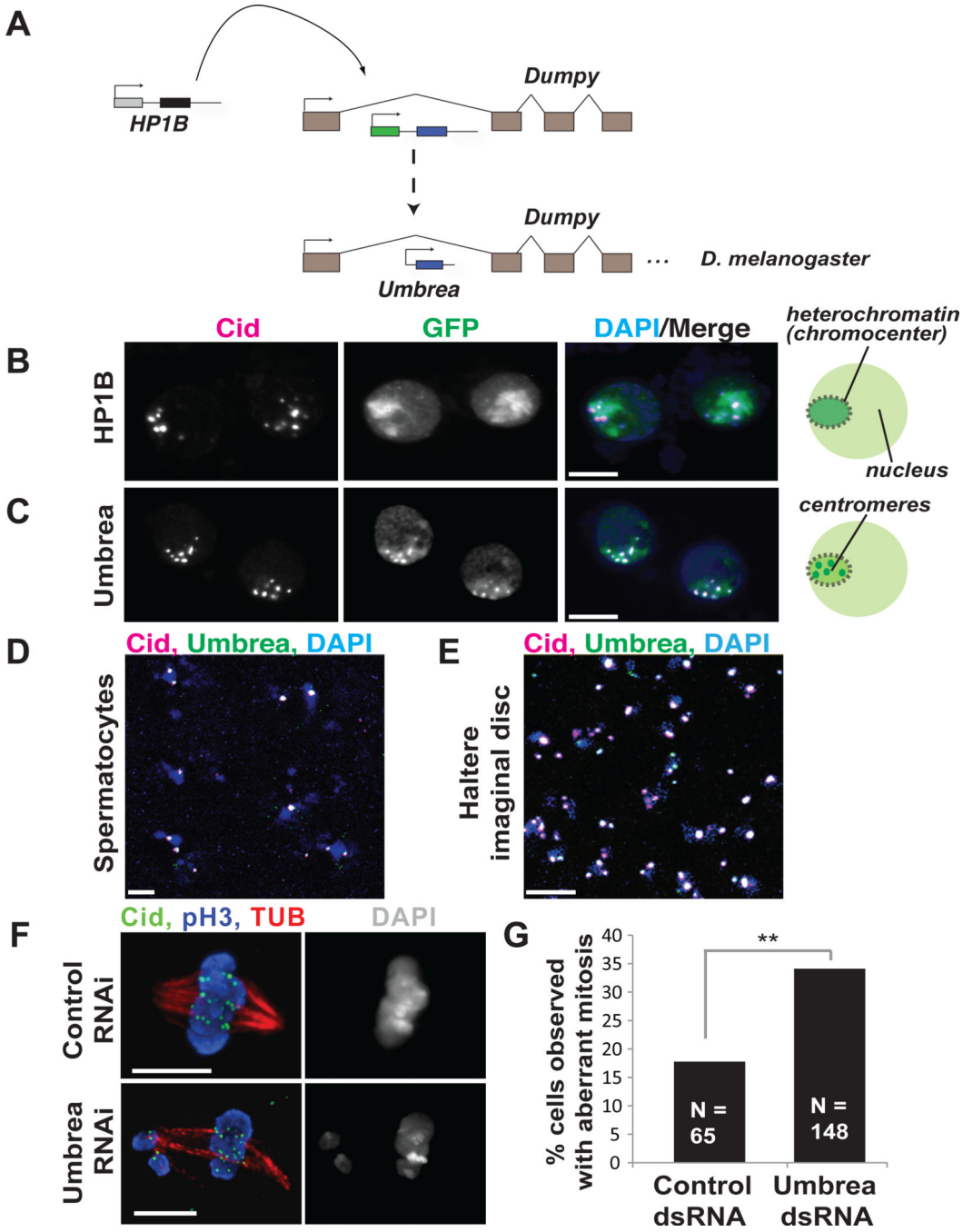

Fig. 1. Neofunctionalization of Umbrea

(A) *Umbrea* originated via gene duplication of the intronless *HP1B* gene into an intron of the *dumpy* locus. **(B)** GFP-tagged HP1B localizes to heterochromatin in *D. melanogaster* Kc cells (GFP in green, anti-Cid in magenta, and DAPI in blue, colocalization appears white). **(C)** In contrast, Umbrea-GFP localizes to centromeres. **(D,E)** Endogenous Umbrea colocalizes with centromeres in testes and in larval imaginal discs (anti-Cid in magenta, anti-Umbrea in green, and DAPI in blue, bar = 5 microns). **(F,G)** S2 cells depleted of

Umbrea by RNAi revealed increased mitotic errors (anti-Cid in green, phospho-H3 staining mitotic chromosomes in blue, and anti-tubulin in red), compared to dsRNA control.

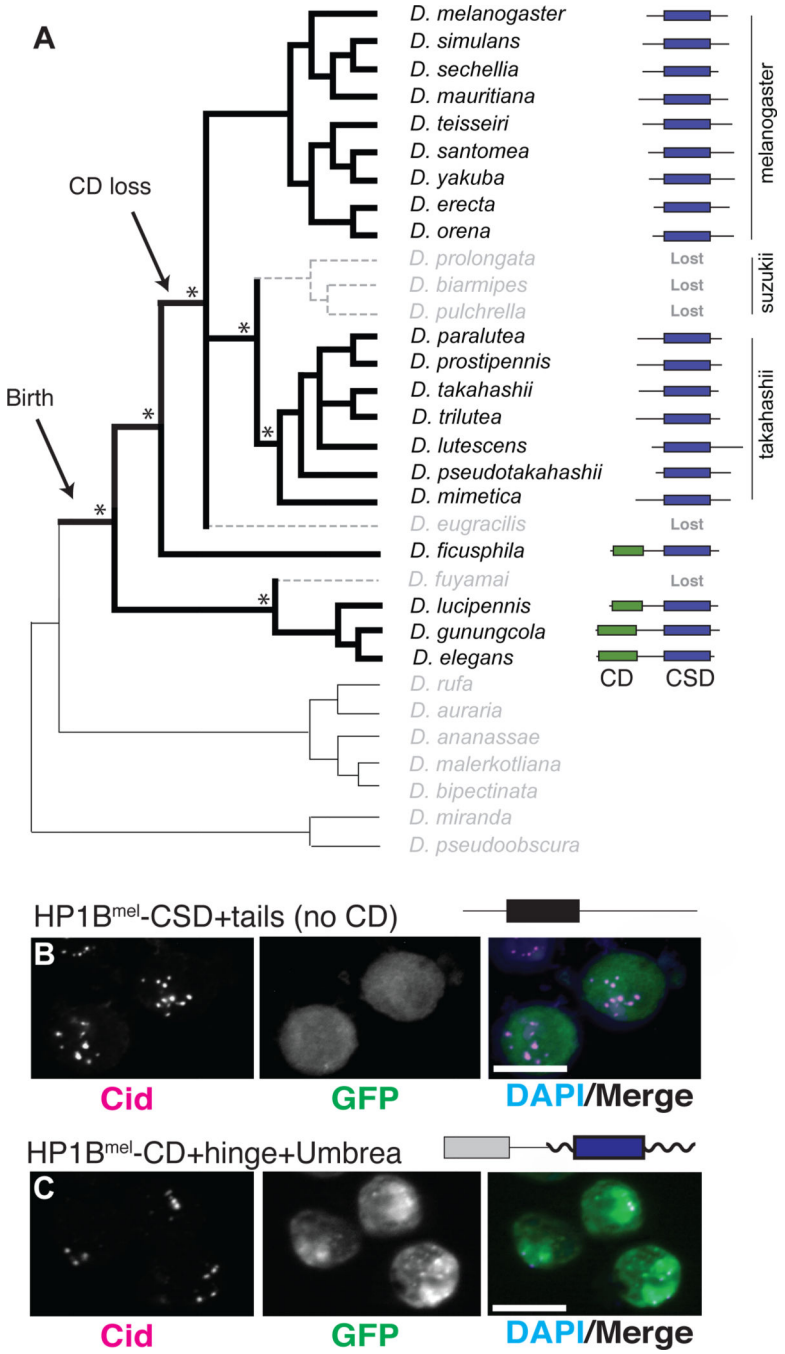

Fig. 2. Dynamic evolution of Umbrea following birth

(A) PCR to shared syntenic sites followed by sequencing (Fig. S6) revealed the presence and structure of *Umbrea* genes. Asterisks indicate strong support for key branch points in the phylogeny(25), suggesting that *Umbrea* was lost at least three times. *Umbrea* is presented with HP1 canonical domains: chromodomain (CD, green) and chromoshadow domain (CSD, blue). **(B)** Localization of GFP-tagged HP1B lacking its CD is diffuse in *D. melanogaster* Kc cell nuclei (GFP is green, anti-Cid is magenta, and DAPI staining of DNA

in blue, bar = 5 microns). **(C)** In contrast, HP1B-CD-hinge fused to Umbrea^{mel} delocalizes it from centromeres.

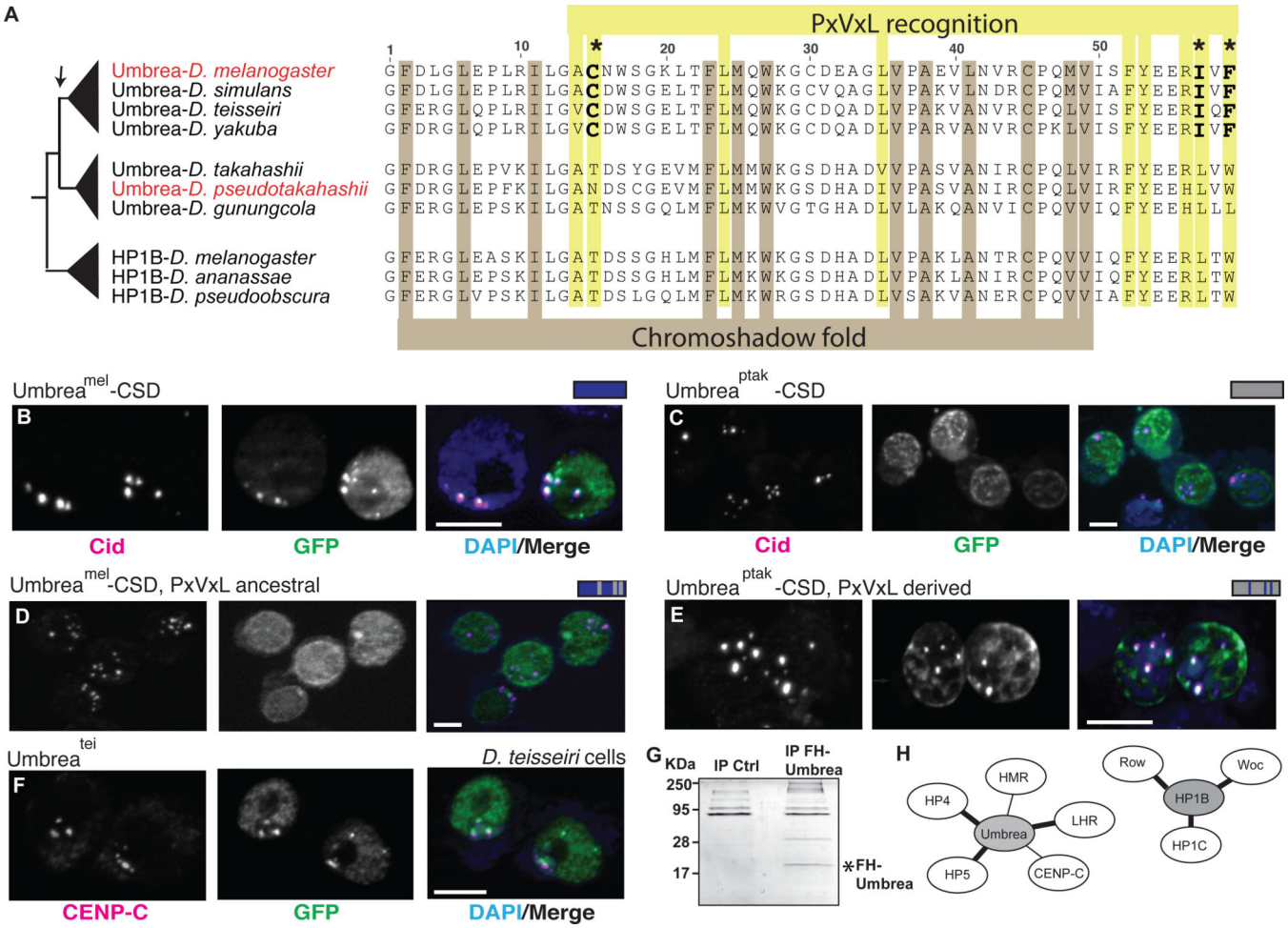

Fig. 3. Chromoshadow changes led to Umbrea centromere localization via altered protein-protein interactions

(A) An amino acid alignment of HP1B and Umbrea CSDs reveals conservation of fold-defining residues but divergence in PxVxL-recognition residues. In particular, 3 changes (bold) are predicted to affect the binding specificity of Umbrea CSD. **(B)** GFP-tagged Umbrea^{mel} CSD (green) colocalizes with Cid (magenta) at centromeres in *D. melanogaster* Kc cells (bar = 5 microns, colocalization appears white). **(C)** However, GFP-tagged Umbrea^{ptak} CSD does not localize to centromeres. **(D)** Reversion of Umbrea^{mel} PxVxL-recognition residues (C-I-F) to ancestral states (T-L-W) causes delocalization from centromeres. **(E)** In contrast, introducing PxVxL-recognition residues (C-I-F) is sufficient to localize Umbrea^{ptak} CSD to centromeres (compare to Fig. 3C). **(F)** Umbrea^{tei} colocalizes with centromeric protein CENP-C in *D. teissieri* cells. **(G)** Immunoprecipitation of Flag-HA-tagged Umbrea pulls down protein complexes in S2 cells. **(H)** Analysis of these complexes reveals that Umbrea and HP1B have mutually exclusive protein-protein interactions. Umbrea interacts with centromere and heterochromatin proteins (Table S2, bold lines indicate confirmation of previously reported interactions(9, 17)), but not with the primary targets of HP1B(18).

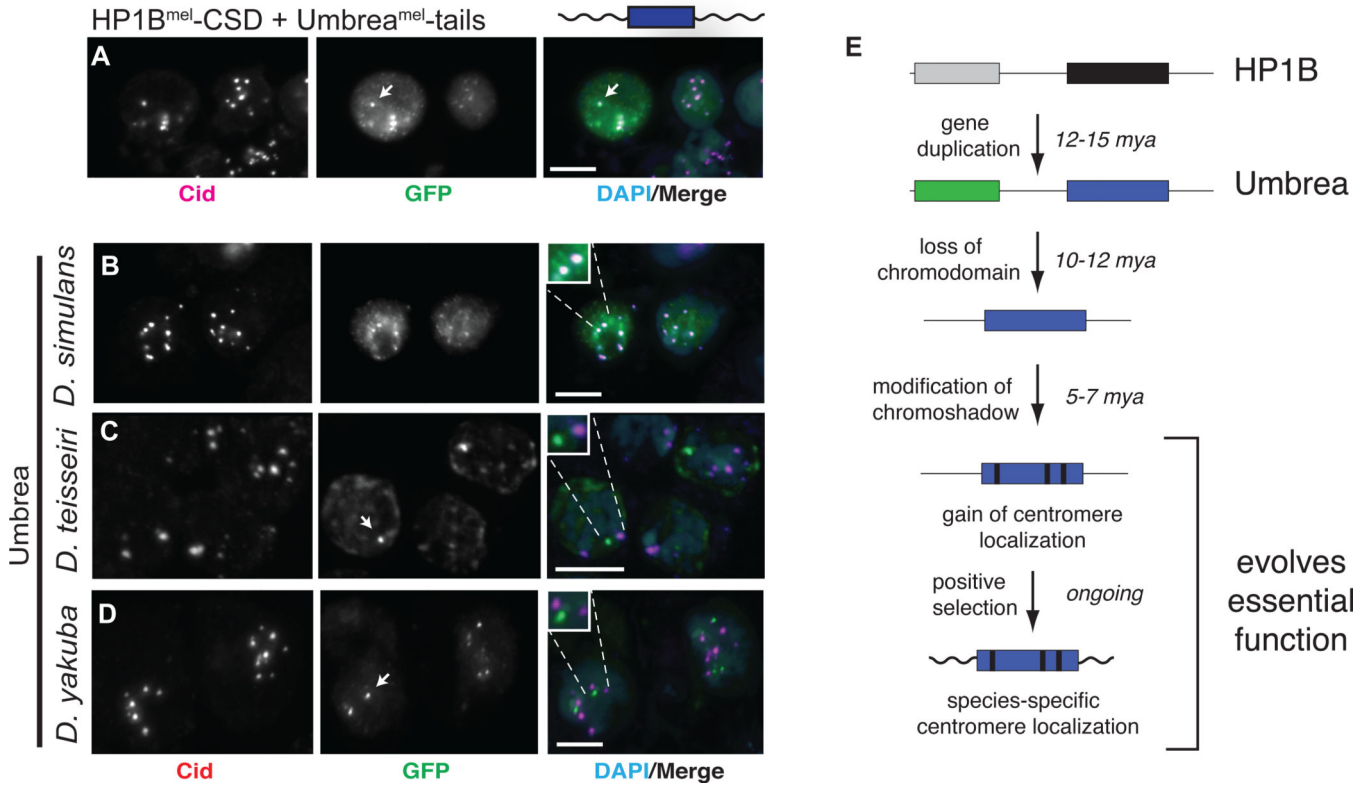

Fig. 4. Species-specific centromere targeting of Umbrea

(A) GFP-fusion of Umbrea^{mel}-tails with HP1B-CD (green) localizes to centromeres (anti-Cid in magenta, bar = 5 microns, colocalization appears white). **(B-D)** *D. melanogaster* Kc cell centromeric localization (Cid, red) of Umbrea orthologs (GFP, green) from *D. simulans, D. teissieri and D. yakuba* worsens with increased divergence. (**E**) Steps to essential neofunctionalization by *Umbrea* following gene duplication.