

15. A. Schubnel, S. Nielsen, J. Taddeucci, S. Vinciguerra, S. Rao, *Earth Planet. Sci. Lett.* **308**, 424 (2011).
16. W. F. Brace, J. D. Byerlee, *Science* **153**, 990 (1966).
17. T. L. Johnson, C. H. Scholz, *J. Geophys. Res.* **81**, 881 (1976).
18. P. G. Okubo, J. H. Dieterich, *J. Geophys. Res.* **89**, 5817 (1984).
19. M. Ohnaka, L.-f. Shen, *J. Geophys. Res.* **104**, 817 (1999).
20. M. Mello, H. S. Bhat, A. J. Rosakis, H. Kanamori, *Tectonophysics* **493**, 297 (2010).
21. E. M. Dunham, R. J. Archuleta, *Geophys. Res. Lett.* **32**, L03302 (2005).
22. T.-F. Wong, *J. Geophys. Res.* **87**, 990 (1982).
23. D. P. Robinson, C. Brough, S. Das, *J. Geophys. Res.* **111**, B08303 (2006).
24. T. Candela, F. Renard, *J. Struct. Geol.* **45**, 87 (2012).
25. R. H. Sibson, *Nature* **316**, 248 (1985).
26. J. R. Rice, *J. Geophys. Res.* **111**, B05311 (2006).
27. G. Di Toro, T. Hirose, S. Nielsen, G. Pennacchioni, T. Shimamoto, *Science* **311**, 647 (2006).
28. N. Brantut, A. Schubnel, J.-N. Rouzaud, F. Brunet, T. Shimamoto, *J. Geophys. Res.* **113**, B10401 (2008).
29. M.-L. Doan, G. Gary, *Nat. Geosci.* **2**, 709 (2009).

Acknowledgments: We thank Y. Piquier for technical support and two reviewers and the associate editor for their constructive remarks, which helped to enhance this paper.

F.X.P., A.S., and S.N. acknowledge support from the French and Italian Ministry of Foreign Affairs program GALILEO, project no. 26019WJ. This work was supported by the Institut National des Sciences de l'Univers. Data are available in the supplementary materials.

Supplementary Materials

www.sciencemag.org/cgi/content/full/340/6137/1208/DC1

Materials and Methods

Figs. S1 to S4

Table S1

References

25 January 2013; accepted 15 April 2013

10.1126/science.1235637

Stepwise Evolution of Essential Centromere Function in a *Drosophila* Neogene

Benjamin D. Ross,^{1,2} Leah Rosin,^{3*} Andreas W. Thoma,^{4*} Mary Alice Hiatt,^{2,5*} Danielle Vermaak,^{2*†} Aida Flor A. de la Cruz,^{2,6} Axel Imhof,⁴ Barbara G. Mellone,³ Harmit S. Malik^{2,6‡}

Evolutionarily young genes that serve essential functions represent a paradox; 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. We traced the evolutionary steps by which the *Drosophila* gene *Umbrea* acquired an essential role in chromosome segregation in *D. melanogaster* since the gene's 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 about how young genes gain essential function. Such innovations may constantly alter the repertoire of centromeric proteins in eukaryotes.

Young essential genes (1) challenge longstanding 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 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.

¹Molecular and Cellular Biology Program, University of Washington, Seattle, WA 98195, USA. ²Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA. ³Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT 06269, USA. ⁴Munich Centre of Integrated Protein Science and Adolf-Butenandt Institute, Ludwig Maximilians University of Munich, 08336 Munich, Germany. ⁵Law Firm of Fitzpatrick, Cella, Harper & Scinto, New York, NY 10104, USA. ⁶Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA.

*These authors contributed equally to this work.

†Present address: Science-wise Consulting, Seattle, WA 98109, USA.

‡Corresponding author. E-mail: hsmalik@fhcrc.org

To gain insight into the birth and evolution of essential function, we focused on one newly evolved gene in *Drosophila*. *Umbrea* (also known as *HP6* and *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 that predominantly localizes to heterochromatin in *D. melanogaster* cells and regulates gene expression (8). *HP1B* is dispensable for viability (8), yet RNA interference (RNAi) knockdown phenotypes show *Umbrea* to be essential in *D. melanogaster* (1, 9). The 100% late larval-pupal lethality upon *Umbrea* knockdown could be rescued by an *Umbrea*-green fluorescent protein (GFP) fusion (fig. S1) further confirmed that *Umbrea* is essential in *D. melanogaster*.

We traced *Umbrea*'s evolutionary path after duplication from *HP1B* to understand when and how essential function was gained by comparing the localization of *HP1B* and *Umbrea* proteins in *D. melanogaster* Kc cells. GFP-tagged *HP1B* proteins from both *D. melanogaster* and *D. ananassae* [whose divergence predates the birth of *Umbrea* (7)] localized to pericentric heterochromatin and euchromatin (Fig. 1B and fig. S2). In contrast, *Umbrea*-GFP predominantly localized to interphase centromeres, but not telomeres (Fig. 1C

and fig. S3, A and B). Specific antibodies raised against *Umbrea* (fig. S4A) confirmed its centromere localization in developing spermatocytes and larval imaginal discs (Fig. 1, D and E, and fig. S4, B and C).

On the basis of its essentiality and centromere localization, we hypothesized that *Umbrea* was required for chromosome segregation. Upon depletion of *Umbrea* by RNAi knockdown (fig. S5A), relative to control cells, *D. melanogaster* S2 cells displayed increased mitotic errors, including delayed chromosome alignment, early anaphase onset, lagging anaphase chromosomes, and multipolar configurations ($P < 0.05$) (Fig. 1, F and G, fig. S5B, and movies S1 to 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 origin of *Umbrea* and subsequent changes, we sequenced the *Umbrea* locus from 32 *Drosophila* species (fig. S6A). Whereas *HP1B* was preserved (7), we found *Umbrea* in only 20 of 32 species, dating its monophyletic origin to 12 to 15 million years ago (Fig. 2A and fig. S6B). Using maximum likelihood methods, we observed evidence of both episodic and recurrent positive selection acting on *Umbrea* (fig. S7, A to 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)—which suggests that *Umbrea* was not essential at or immediately after its birth.

Four lineages retained full-length *Umbrea* genes, two of which encode an intact chromodomain (CD) and ancestral residues essential for binding histone H3 trimethyl Lys⁹ (H3K9me) (fig. S8) (11). However, most extant *Umbrea* genes have lost their CDs, and encode 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), which suggests 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

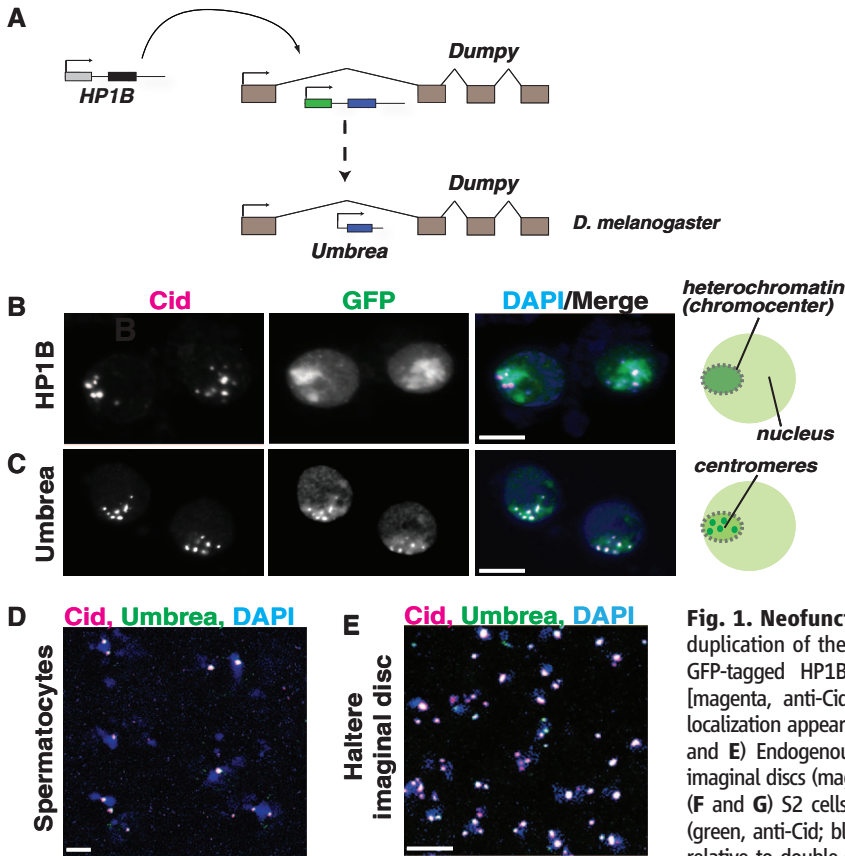


Fig. 1. Neofunctionalization of Umbrea. (A) *Umbrea* originated via gene duplication of the intronless *HP1B* gene into an intron of the *dumpy* locus. (B) GFP-tagged *HP1B* localizes to heterochromatin in *D. melanogaster* Kc cells [magenta, anti-Cid; green, GFP; blue, 4',6-diamidino-2-phenylindole (DAPI); colocalization appears white]. (C) In contrast, *Umbrea*-GFP localizes to centromeres. (D and E) Endogenous *Umbrea* colocalizes with centromeres in testes and in larval imaginal discs (magenta, anti-Cid; green, anti-*Umbrea*; blue, DAPI; scale bar, 5 μm). (F and G) S2 cells depleted of *Umbrea* by RNAi revealed increased mitotic errors (green, anti-Cid; blue, phospho-H3-staining mitotic chromosomes; red, anti-tubulin) relative to double-stranded RNA control (***P* < 0.05).

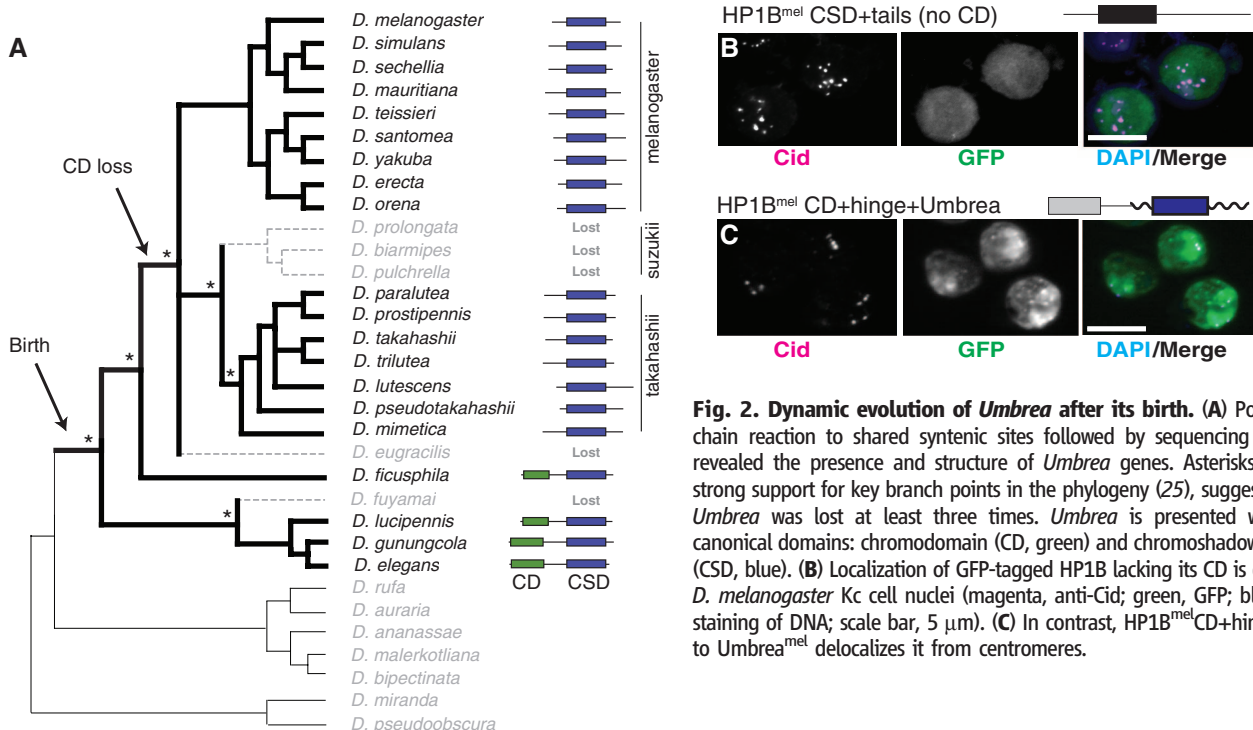


Fig. 2. Dynamic evolution of Umbrea after its birth. (A) Polymerase chain reaction to shared syntenic sites followed by sequencing (fig. S6) revealed the presence and structure of *Umbrea* genes. Asterisks indicate strong support for key branch points in the phylogeny (25), suggesting that *Umbrea* was 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 (magenta, anti-Cid; green, GFP; blue, DAPI staining of DNA; scale bar, 5 μm). (C) In contrast, *HP1B*^{mel}CD+hinge fused to *Umbrea*^{mel} delocalizes it from centromeres.

degenerate PxVxL motifs (P, Pro; V, Val; L, Leu; x, any amino acid) (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 and fig. S9). We found that *D. melanogaster* Umbrea CSD localized to centromeres (Fig. 3B). This property was not shared with HP1B CSD or even other Umbrea CSDs, because neither “parental” HP1B^{mel} CSD (from *D. melanogaster*) nor Umbrea^{ptak} CSD (from *D. pseudotakahashii*) could localize to centromeric regions in *D. melanogaster* cells (Fig. 3B and fig. S10B). We conclude that a discrete transition for centromere localization occurred in Umbrea CSD after divergence of the melanogaster and takahashii subgroups, co-

incident with changes in the PxVxL recognition residues. Indeed, reversion of these three residues (Cys¹⁵, Ile⁵⁷, and Phe⁵⁹; Fig. 3A and fig. 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 centromere localization by Umbrea CSD originated in the common ancestor of the melanogaster species subgroup 5 to 7 million years ago. Consistent with this, we found that GFP-Umbrea^{tei} localized 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 local-

ization by alteration of protein interactions, we performed proteomic analyses to identify proteins that coimmunoprecipitate 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. We found no overlap with protein partners of HP1B, which include the euchromatic proteins HP1C, Woc, and Row (18) (Fig. 3H); this suggests a rewiring of the protein interaction network of Umbrea.

Our evolutionary analyses (fig. S7, A to 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. HP1B^{mel}

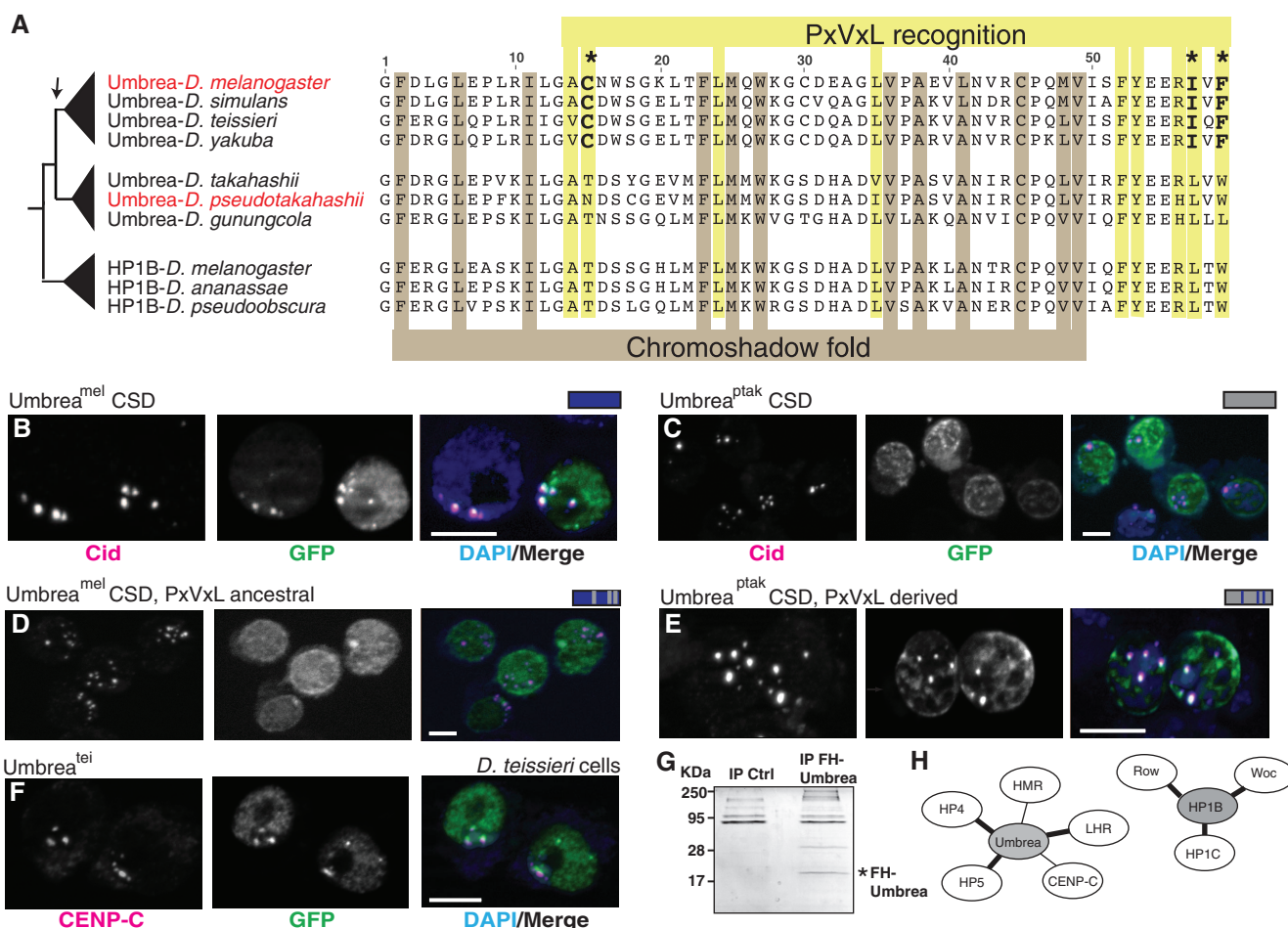
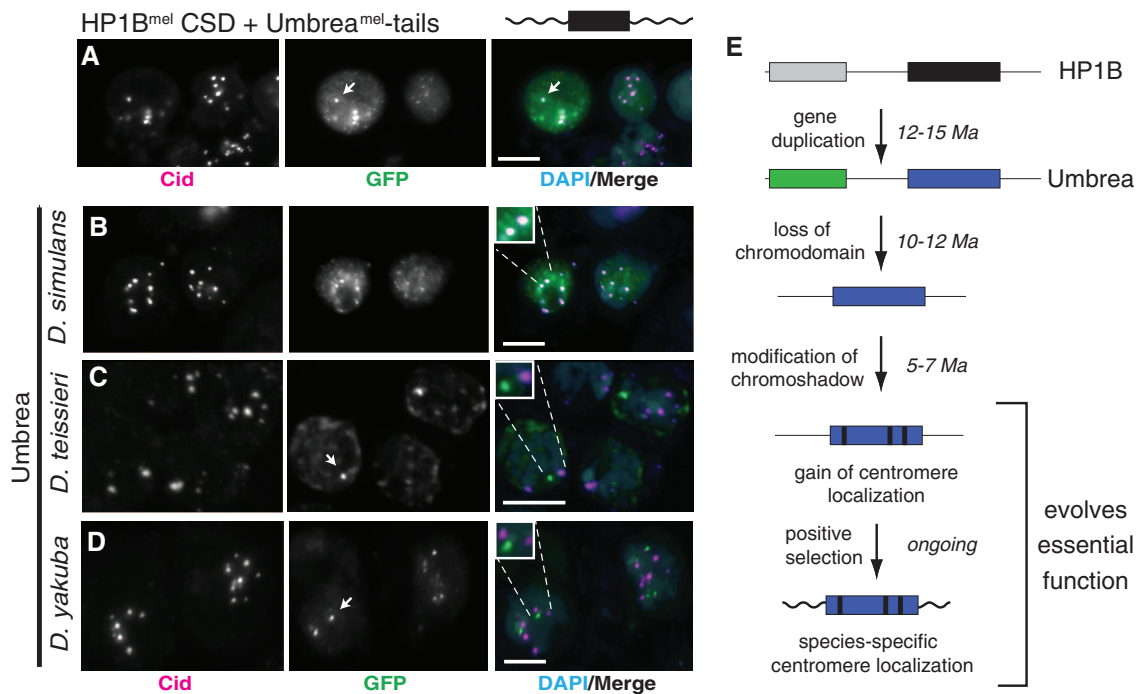


Fig. 3. Chromoshadow changes led to Umbrea centromere localization via altered protein-protein interactions. (A) An amino acid alignment of HP1B and Umbrea CSDs reveals conservation of fold-defining residues but divergence in PxVxL recognition residues. In particular, three changes (bold) are predicted to affect the binding specificity of Umbrea CSD. Abbreviations for amino acid residues: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr. (B) GFP-tagged Umbrea^{mel} CSD (green) colocalizes with Cid (magenta) at centromeres in *D. melanogaster* Kc cells (scale bar, 5 μm, colocalization appears white). (C) However, GFP-tagged Umbrea^{ptak} CSD does not localize to centromeres. (D) Reversion of

Umbrea^{mel} PxVxL recognition residues (Cys-Ile-Phe) to ancestral states (Thr-Leu-Trp) causes delocalization from centromeres. (E) By contrast, introduction of PxVxL recognition residues (Cys-Ile-Phe) 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 (F)- and hemagglutinin (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 (magenta, anti-Cid; colocalization appears white; scale bar, 5 μ m). (B to D) *D. melanogaster* Kc cell centromere localization (magenta, Cid) of *Umbrea* orthologs (green, GFP) from *D. simulans*, *D. teissieri*, and *D. yakuba* worsens with increased divergence. (E) Steps to essential neofunctionalization by *Umbrea* after gene duplication (Ma, millions of years ago).



CSD alone showed no discrete localization (Fig. 2B), whereas the 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. Whereas the CSD likely mediates its localization via protein-protein interactions, *Umbrea* tails may bind centromeric nucleic acids, analogous to the hinge region of mammalian HP1 α , which binds DNA in vitro (19). Because 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. 4, C and D), localizing instead to distinct foci. Although positive selection of *Umbrea* preceded its centromere 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; 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 centromeric proteins (22), we propose that recurrent changes at centromeric DNA satellites could drive the retention of duplicate genes such as *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 before 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 among centromeric proteins across taxa (24). Although 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.

References and Notes

1. S. Chen, Y. E. Zhang, M. Long, *Science* **330**, 1682 (2010).
2. G. L. Miklos, G. M. Rubin, *Cell* **86**, 521 (1996).
3. A. Force et al., *Genetics* **151**, 1531 (1999).
4. A. Stoltzfus, *J. Mol. Evol.* **49**, 169 (1999).
5. H. Innan, F. Kondrashov, *Nat. Rev. Genet.* **11**, 97 (2010).
6. J. Zhang, A. M. Dean, F. Brunet, M. Long, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 16246 (2004).
7. M. T. Levine et al., *PLoS Genet.* **8**, e1002729 (2012).
8. D. Zhang, D. Wang, F. Sun, *Chromosoma* **120**, 97 (2011).
9. C. Joppich, S. Scholz, G. Korge, A. Schwendemann, *Chromosome Res.* **17**, 19 (2009).

10. M. W. Hahn, *J. Hered.* **100**, 605 (2009).
11. S. A. Jacobs, S. Khorasanizadeh, *Science* **295**, 2080 (2002).
12. N. Murzina, A. Verreault, E. Laue, B. Stillman, *Mol. Cell* **4**, 529 (1999).
13. J. Nakayama, J. C. Rice, B. D. Strahl, C. D. Allis, S. I. Grewal, *Science* **292**, 110 (2001).
14. T. Ohta, *Genetics* **120**, 841 (1988).
15. J. F. Smothers, S. Henikoff, *Curr. Biol.* **10**, 27 (2000).
16. A. Thiru et al., *EMBO J.* **23**, 489 (2004).
17. L. Giot et al., *Science* **302**, 1727 (2003).
18. J. Abel, R. Eskeland, G. D. Raffa, E. Kremmer, A. Imhof, *PLoS ONE* **4**, e5089 (2009).
19. R. R. Meehan, C. F. Kao, S. Pennings, *EMBO J.* **22**, 3164 (2003).
20. A. R. Lohe, D. L. Brutlag, *J. Mol. Biol.* **194**, 161 (1987).
21. D. Vermaak, H. S. Hayden, S. Henikoff, *Mol. Cell. Biol.* **22**, 7553 (2002).
22. H. S. Malik, S. Henikoff, *Cell* **138**, 1067 (2009).
23. M. D. Daugherty, H. S. Malik, *Annu. Rev. Genet.* **46**, 677 (2012).
24. I. M. Cheeseman, A. Desai, *Nat. Rev. Mol. Cell Biol.* **9**, 33 (2008).
25. B. Prud'homme et al., *Nature* **440**, 1050 (2006).

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. Supported by NIH Training grant T32HG000035 and an NSF predoctoral fellowship (B.D.R.), an EU network grant (EpiGeneSys 257082, A.I.), NSF award 1024973 (B.G.M.), and a grant from the Mathers Foundation and NIH grant R01GM074108 (H.S.M.). H.S.M. is an HHMI Early Career Scientist. *Umbrea* DNA sequences have been submitted to GenBank under accession numbers KC660086 to KC660100. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (PRIDE partner repository dataset identifier PXD000163).

Supplementary Materials

www.sciencemag.org/cgi/content/full/340/6137/1211/DC1
Materials and Methods
Figs. S1 to S10
Table S1
References (26–38)
Movies S1 to S3

21 December 2012; accepted 1 April 2013
10.1126/science.1234393

Stepwise Evolution of Essential Centromere Function in a *Drosophila* Neogene

Benjamin D. Ross, Leah Rosin, Andreas W. Thomae, Mary Alice Hiatt, Danielle Vermaak, Aida Flor A. de la Cruz, Axel Imhof, Barbara G. Mellone and Harmit S. Malik

Science **340** (6137), 1211-1214.
DOI: 10.1126/science.1234393

Essential Novelty

The evolution of essential function for newly originated genes presents a conundrum, in that prior to the gene's origin either the essential function was absent or else performed by another gene or set of genes. In order to better understand how new genes acquire essential function, **Ross et al.** (p. 1211) investigated the origin of the *Drosophila* gene *Umbrea*. *Umbrea* became an essential protein in certain *Drosophila* species through the gain of localization at the centromere and a role in chromosome segregation.

ARTICLE TOOLS

<http://science.sciencemag.org/content/340/6137/1211>

SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2013/06/05/340.6137.1211.DC1>

REFERENCES

This article cites 38 articles, 12 of which you can access for free
<http://science.sciencemag.org/content/340/6137/1211#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)