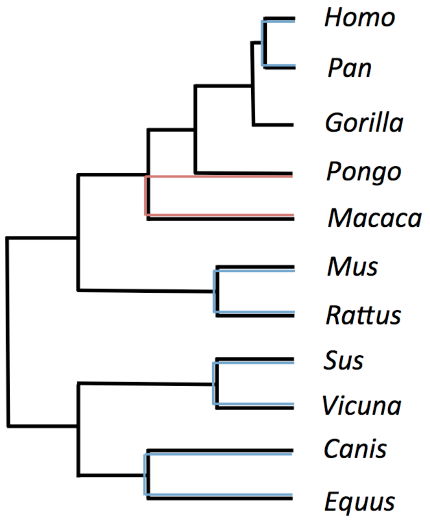
Supplemental Methods. Application of a Total Least Squares approach to dating divergence times in mammalian and Drosophila clades.

We applied this novel approach to dating divergence times to two other taxonomic groups with extensive genomics resources, in particular placental mammals and *Drosophila*.

Analysis of the mammalian datasest

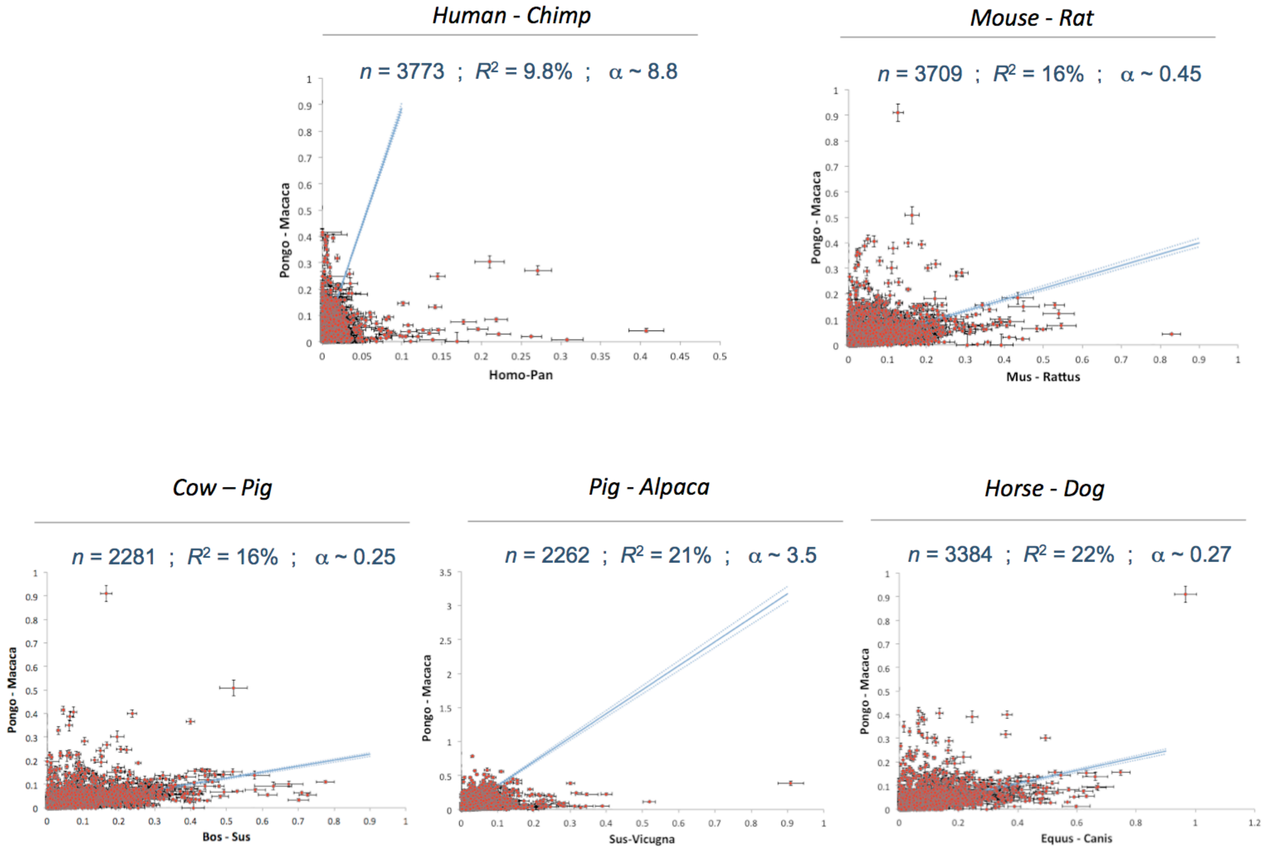
For the mammalian analyses we used the dataset of coding sequences available at "OrthoMam: a Database of Mammalian Markers" (Douzery, et al. 2014). We chose a few representative species of each mammalian order from the total dataset, and used the pair *Pongo*-*Macaca* as the reference (constant) species pair to which all other pairs were compared (Suppl. Fig. S1).



Supplemental Figure S1. Phylogenetic relationships among taxa used in the pairwise comparisons among placental mammals shown in Supp. Fig. S2 (Douzery, et al. 2014). The *Pongo – Macaca* pair (red) was used as the comparator against which all other taxon pairs (shown in blue) were

Data generation: the OrthoMam database (http://www.orthomam.univ-montp2.fr/orthomam/) contains sets of curated and aligned orthologous genes for mammalian species. At the time of the analysis, the database contained alignments for close to 4000 protein-coding genes. Similarly to the approach used for *Plasmodium* genes, for each four-species comparison we identified all genes with homologs present in all four species, and for which the length of the shortest of the four sequences was at least 90% the length of the longest (see Materials and Methods). Amino acid sequence divergence was estimated in described (Materials and Methods). The size of the dataset, *n*, in each comparison is shown in Suppl. Fig. S2.

In a stark contrast to the *Plasmodium* results, in each of the four-way mammalian comparisons, the TLS regression explained only a small fraction of the variance in the data, with R^2 values ranging from 10% to 22% (Suppl. Fig. S2). Since the evolution of mammalian proteins does not conform well to a simple molecular clock model, the approach to dating the relative age of lineages based on the slope of the TLS is not valid (see main text and Fig. 1).


Human - Chimp

$n = 3773$; $R^2 = 9.8\%$; $\alpha \sim 8.8$

Mouse - Rat

$n = 3709$; $R^2 = 16\%$; $\alpha \sim 0.45$

Cow – Pig

$n = 2281$; $R^2 = 16\%$; $\alpha \sim 0.25$

Pig - Alpaca

$n = 2262$; $R^2 = 21\%$; $\alpha \sim 3.5$

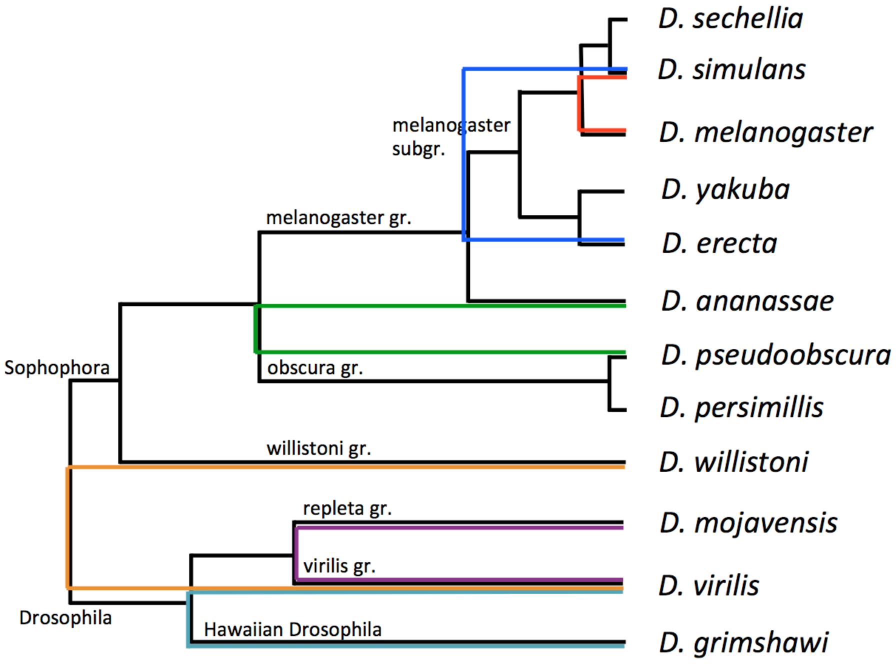
Horse - Dog

$n = 3384$; $R^2 = 22\%$; $\alpha \sim 0.27$

Supplemental Figure S2. Relationship of amino acid sequence divergence between independent mammalian lineages (see Suppl. Fig. S1 above). In each plot, each point represents a protein present in the four taxa compared. The TLS regression fit (solid line) and 95% confidence intervals (dotted lines) are shown in blue; the slope of the line, α, equals the ratio between the divergence time of each of five species pairs represented in the X axis relative to that of *Pongo*-*Macaca*, which was used as the common pair in all comparisons (Y axis).

Analysis of the *Drosophila* datasest

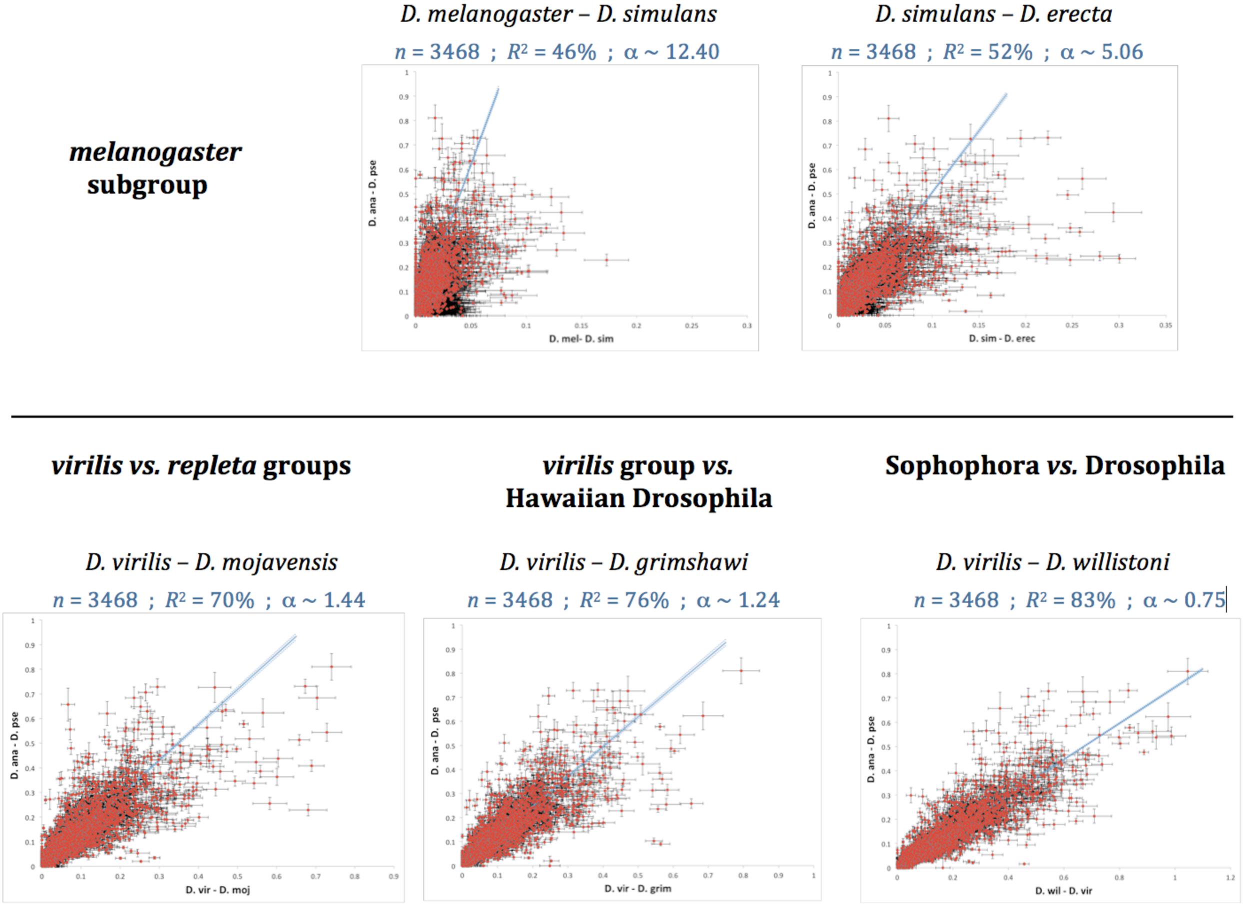
The *Drosophila* dataset was obtained from the *Drosophila* 12 Genomes project (Drosophila 12 Genomes, et al. 2007). We chose representatives of several major cladogenesis events within the genus *Drosophila*, and used as the common comparator pair two species that represent the split between the *melanogaster* (*D. ananassae*) and the *obscura* (*D. pseudoobscura*) species groups (Suppl. Fig. S3). We used the curated 12-species protein alignments available through FlyBase (http://flybase.org/).



Supplemental Figure S3. Phylogeny of *Drosophila* species, including representatives of the *Drosophila* and *Sophophora* subgenera (Drosophila 12 Genomes, et al. 2007). The divergence between the *melanogaster* and the *obscura* species groups (green), represented, respectively, by *D. ananassae* and *D.*

pseudoobscura, was compared with the divergence of five other speciation events: a) the split between *D. melanogaster* and *D. simulans* (red), (b) the origin of the melanogaster subgroup, represented by the split *D. erecta*-*D. simulans* (blue), (c) the split between the *repleta* and *virilis* species groups, represented by *D. mojavensis* and *D. virilis*, respectively (violet), (d) the split between the subgenera Sophophora (*D. willistoni*) and Drosophila (*D. virilis*) (orange), and (e) the split between the Hawaiian *Drosophila* (*D. grimshawi*) and the *repleta/virilis* lineage (*D. virilis*) (teal).

Data generation: FlyBase contains sets of curated and aligned orthologous genes for the twelve *Drosophila* species. The dataset was selected as described for *Plasmodium* and mammals (see Materials and Methods); 3468 genes passed our selection criteria in all 4-species comparisons and were used in the analyses (Suppl. Fig. S4).


***melanogaster* subgroup**

D. melanogaster – D. simulans

n = 3468 ; *R*² = 46% ; α ~ 12.40

D. simulans – D. erecta

n = 3468 ; *R*² = 52% ; α ~ 5.06

***virilis vs. repleta* groups**

D. virilis – D. mojavensis

n = 3468 ; *R*² = 70% ; α ~ 1.44

***virilis* group *vs.* Hawaiian Drosophila**

D. virilis – D. grimshawi

n = 3468 ; *R*² = 76% ; α ~ 1.24

Sophophora *vs.* Drosophila

D. virilis – D. willistoni

n = 3468 ; *R*² = 83% ; α ~ 0.75

Supplemental Figure S4. Relationship of amino acid sequence divergence between independent *Drosophila* lineages (described in Suppl. Fig. S3). Legend as in Suppl. Fig. S2. The reference species pair is *D. ananassae-D.pseudoobscura* (D. ana-D. pse), shown in the Y axis.

Drosophila proteins, much like those in *Plasmodium*, conform very nicely to the molecular clock model, with R^2 that ranged between 46% and 83% in the five comparisons (Suppl. Fig. S4). Therefore, we used the slope of each regression, α, as an estimator of the relative age of the two lineages being compared.

Supplemental Table S4. Divergence time estimates for *Drosophila* clades.

Clades[1]	**Species comparison**	**α**[2]	**Age (MY)**[3]	**Age from α (MY)**[4]
melanogaster – obscura groups	*D.ana – D.pseudo*		24 (17-31)	
D. melanogaster – D. simulans	*D.mel – D.sim*	0.08	1.4 (0.9-1.9)	1.9
melanogaster subgroup	*D.sim – D.erec*	0.20	3.3 (2.4-4.4)	4.8
virilis – repleta groups	*D.vir – D.moj*	0.70	10 (6.5-13)	16.8
virilis group – Hawaiian Drosophila	*D.vir – D.grim*	0.81	13 (9.6-17)	19.4
Sophophora – Drosophila	*D.will – D.vir*	1.34	32 (25-40)	32.2

[1] Clades colored as in Suppl. Fig. S3.

[2] Alpha, α, obtained from TLS regression (Suppl. Fig. S4).

[3] Age estimates as in Obbard et al. (Obbard, et al. 2012), in million years (MY).

[4] Divergence of the *melanogaster* and *obscura* assumed to have occurred 24 MY ago (Obbard, et al. 2012). Dates are obtained by multiplying 24 MY by α.

The age of each clade relative to the age of the speciation event that gave rise to the *melanogaster* and *obscura* species groups 24 million years (MY) ago was estimated based on α (Suppl. Table S4; last column). These dates are very similar to, and often towards the upper end of, those obtained based on molecular data (Obbard, et al. 2012), and well in line with fossil evidence (Grimaldi 1987, 1988).

References

Douzery EJ, Scornavacca C, Romiguier J, Belkhir K, Galtier N, Delsuc F, Ranwez V. 2014. OrthoMaM v8: A Database of Orthologous Exons and Coding Sequences for Comparative Genomics in Mammals. Mol Biol Evol.

Drosophila 12 Genomes C, Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W, et al. 2007. Evolution of genes and genomes on the Drosophila phylogeny. Nature 450:203-218.

Grimaldi DA. 1987. Amber fossil Drosphilidae (Diptera), with particular reference to the Hispaniolan taxa. American Museum Novitates 2880:1-23.

Grimaldi DA. 1988. Relicts in the Drosophilidae (Diptera). In: Liebherr JK, editor. Zoogeography of Caribbean insects. Ithaca (NY): Cornell University Press. p. 183-213.

Obbard DJ, Maclennan J, Kim KW, Rambaut A, O'Grady PM, Jiggins FM. 2012. Estimating divergence dates and substitution rates in the Drosophila phylogeny. Mol Biol Evol 29:3459-3473.