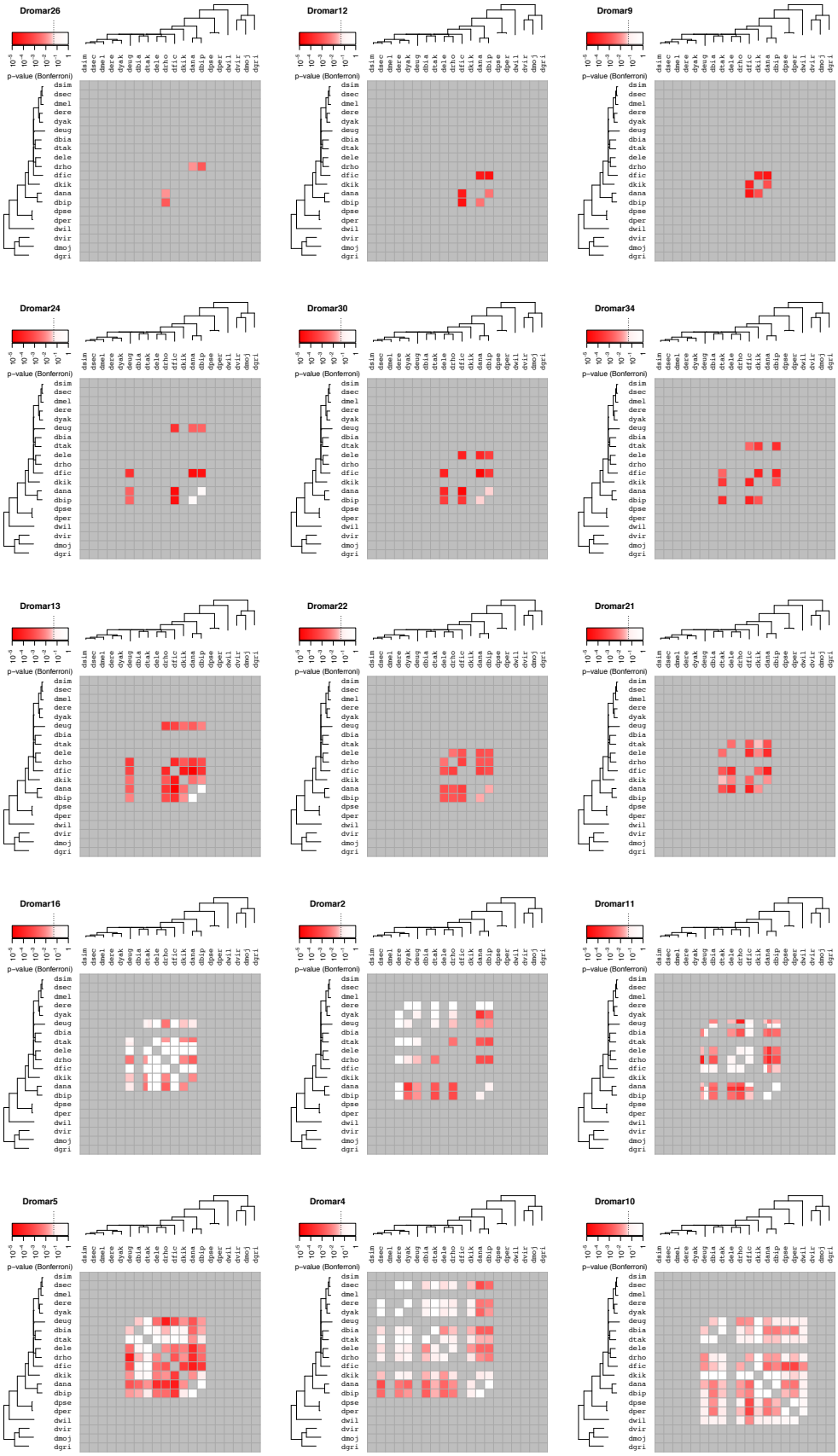
Table 1: Single copy orthologous genes from the 20 *Drosophila* genomes used in the dS estimate.

Single Copy Orthologous Host Gene – Flybase Name	Full name	Chromossome location	FlyBaseID
Amd	α methyl dopa	2L	FBgn0000075
Ddc	dopa decarboxylase	2L	FBgn0000422
Mp20	muscle protein 20	2R	FBgn0002789
Adh	alcohol decarboxylase	2L	FBgn0000056
AnnX-RA	Annexin X	X	FBgn0000084
Rab35	Rab35	X	FBgn0031090
pickel	pickel	X	FBgn0013720
Adar	Adenosine deaminase acting on RNA	X	FBgn0026086
Pgd	Phosphogluconate dehydrogenase	X	FBgn0004654
kirre	kin of irre	X	FBgn0024980
Mcm3	Minichromosome maintenance 3	X	FBgn0024332
CG42265	CG42265	X	FBgn0259150
CG4593	CG4593	X	FBgn0029929
CG14434	CG14434	X	FBgn0029915
CG14435	CG14435	X	FBgn0029911
Gclc	Glutamate-cysteine ligase catalytic subunit	X	FBgn0040319
CG10959	CG10959	X	FBgn0030010
CG12065	CG12065	X	FBgn0030052
Lim1	Lim1	X	FBgn0026411
CG42388	CG42388	X	FBgn0259734
CG7033	CG7033	X	FBgn0030086
AP-1γ	AP-1γ	X	FBgn0030089
Bx42	Bx42	X	FBgn0004856
CG1986	CG1986	X	FBgn0030162
grau	grauzone	2R	FBgn0001133
CG42381	CG42381	2R	FBgn0259727
CG4386	CG4386	2R	FBgn0034661
GlcT-1	GlcT-1	2R	FBgn0067102
Vps35	Vacuolar protein sorting 35	2R	FBgn0034708
CG2921	CG2921	2R	FBgn0034689
wrapper	wrapper	2R	FBgn0025878
CG4610	CG4610	2R	FBgn0034735
Vps20	Vacuolar protein sorting 20	2R	FBgn0034744

blw	bellwether	2R	FBgn0011211
Gmer	GDP-4-keto-6-deoxy-D-mannose 3,5-epimerase/4-reductase	2R	FBgn0267823
CG30196	CG30196	2R	FBgn0050196
Art7	Arginine methyltransferase 7	2R	FBgn0034817
CG4091	CG4091	2R	FBgn0034894
CG2812	CG2812	2R	FBgn0034931
CG3209	CG3209	2R	FBgn0034971
Ca-P60A-RA	Ca-P60A-RA	2R	FBgn0263006
Nap1	Nucleosome assembly protein 1	2R	FBgn0015268
CG4585	CG4585	2R	FBgn0025335
CG4797	CG4797	2R	FBgn0034909
CG3419	CG3419	2R	FBgn0266438
CG16912	CG16912	2R	FBgn0035064
cln3	cln3	3L	FBgn0036756
PRL-1-RB	PRL-1-RB	2L	FBgn0024734
TER94-RA	TER94-RA	2R	FBgn0261014
Vps25-RA	Vps25-RA	2R	FBgn0022027
abd-A	abdominal A	3R	FBgn0000014
Abl	Abl tyrosine kinase	3L	FBgn0000017
ac	achaete	X	FBgn0000022
CG9586	CG9586	2L	FBgn0032101
vnc	variable nurse cells	3L	FBgn0263251
Csat	Csat	X	FBgn0024994
Acyp	Acylphosphatase	2L	FBgn0025115
al	aristaless	2L	FBgn0000061
aop	anterior open	2L	FBgn0000097
b	black	2L	FBgn0000153
bib	big brain	2L	FBgn0000180
bur	burgundy	2L	FBgn0000239
cact	cactus	2L	FBgn0000250
nAChRβ1	nicotinic Acetylcholine Receptor β1	3L	FBgn0000038
Aprt	Adenine phosphoribosyltransferase	3L	FBgn0000109
D	Dichaete	3L	FBgn0000411
dib	disembodied	3L	FBgn0000449
ect	ectodermal	3L	FBgn0000451
Pka-C3	Protein kinase, cAMP-dependent, catalytic subunit 3	3L	FBgn0000489
E(z)	Enhancer of zeste	3L	FBgn0000629

p130CAS	p130CAS	3L	FBgn0035101
Vdup1	Vitamin D3 up-regulated protein 1	3L	FBgn0035103
rno	rhinoceros	3L	FBgn0035106
pyx	pyrexia	3L	FBgn0035113
dbo	diablo	3L	FBgn0040230
bc10	bc10	3L	FBgn0040239
DNApol-α73	DNA polymerase α 73kD	3R	FBgn0005696
Eip63E	Ecdysone-induced protein 63E	3L	FBgn0005640
ple	pale	3L	FBgn0005626
CCAP	Crustacean cardioactive peptide	3R	FBgn0039007
Usp12-46	Ubiquitin-specific protease 12/46 ortholog	3R	FBgn0039025
Lis-1	Lissencephaly-1	2R	FBgn0015754
Cnx99A	Calnexin 99A	3R	FBgn0015622
AstA	Allatostatin A	3R	FBgn0015591
Apc	APC-like	3R	FBgn0015589
sda	slamdance	3R	FBgn0015541
spel1	spellchecker1	2L	FBgn0015546
RpS21	Ribosomal protein S21	2L	FBgn0015521
hoip	hoi-polloi	2L	FBgn0015393
cutlet	cutlet	2L	FBgn0015376
CG5861	CG5861	2L	FBgn0015338
bap	bagpipe	3R	FBgn0004862
C15	C15	3R	FBgn0004863
cdi	center divider	3R	FBgn0004876
tok	tolkin	3R	FBgn0004885
tws	twins	3R	FBgn0004889
Arl2	ADP ribosylation factor-like 2	3R	FBgn0004908
Gnf1	Germ line transcription factor 1	3R	FBgn0004913
Calr	Calreticulin	3R	FBgn0005585
EloB	Elongin B	3R	FBgn0023212

Figure S1: VHICA p-value matrix for 15 *mariner* elements in the 20 *Drosophila* genome.
Gray squares: no comparison available



Supplementary material S3

Evolutionary correlations of Codon Usage Bias

In the VHICA method, we used the average Effective Number of Codons (ENC) between two species to account for the strength of selection on synonymous substitutions. By taking the average ENC, we do not assume that ENC are identical between species, but rather than the average ENC reflects the average evolutionary selection strength on sequences during species divergence.

Figure S3.1 shows that ENCs are very correlated between close species, but this correlation decreases (down to 0.3) for the most distant species. There is thus a global conservation signal of ENC across the Drosophila genus, but the ENC is clearly phylogenetically correlated.

Figure S3.2 shows that the choice of ENC measurement used to compute the P-values of horizontal transfers is unlikely to affect the results qualitatively. For this pair of species (*D. bipectinata* and *D. rhopaola*), all significant HTTs remain significant even when taking the grand mean ENC or the ENC of each species instead of the default method.

Figure S3.3 illustrates the relationships between ENC and dS for different pairs of species. For close species (e.g. *D. ananassae* and *D. bipectinata*), ENCs are largely correlated (and dS scores are low). When the distance between species increases, the correlation between ENCs decreases as well. *D. willistoni* displays a unique pattern among the 20 species, with a very low CUB (high ENC) for most genes. This does not harm the VHICA analysis, as the correlation between ENC and dS is still present (small symbols on the bottom left , large symbols on the top right part of the figure).

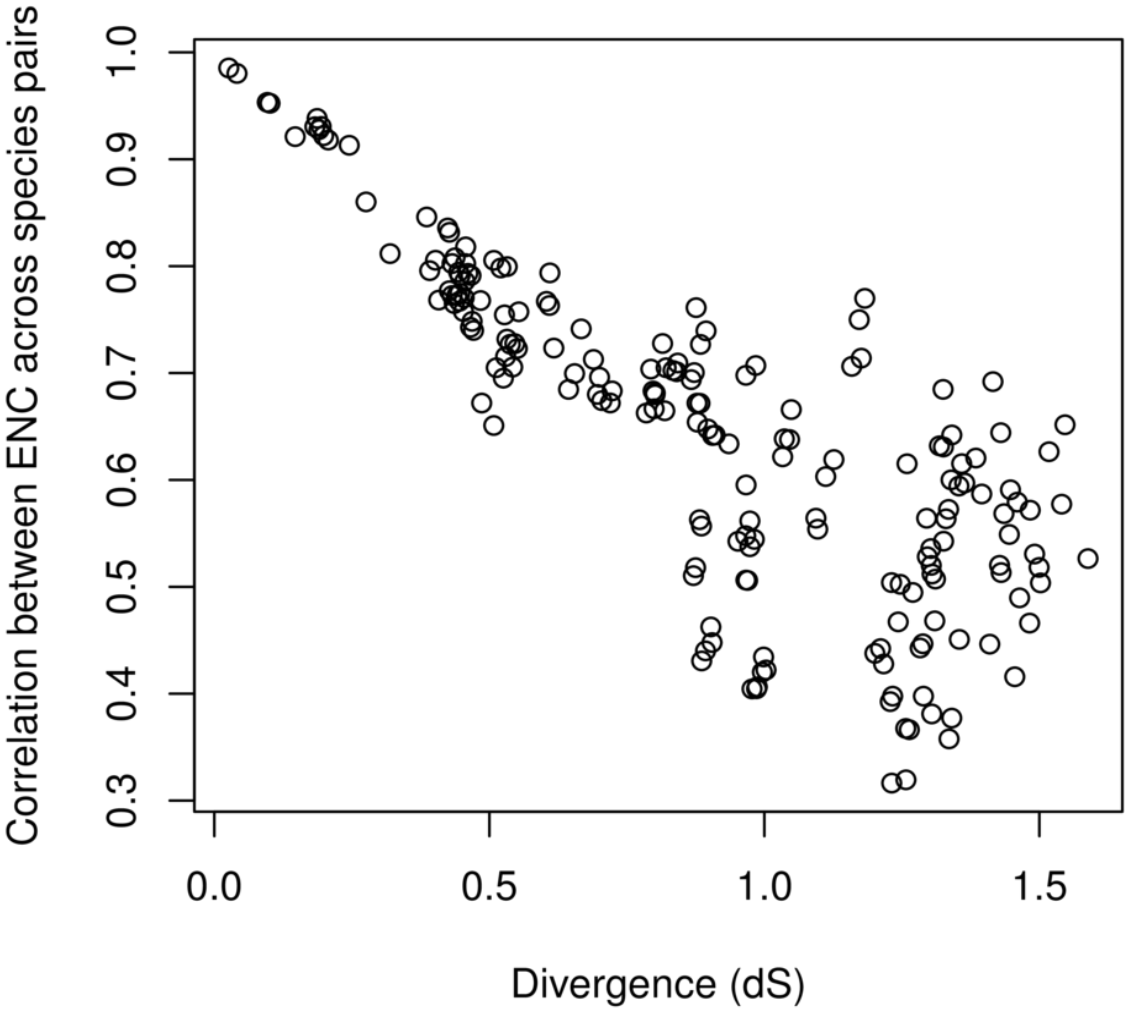


Figure 1: Correlation between the ENC of 50 genes as a function of the distance between species pairs.

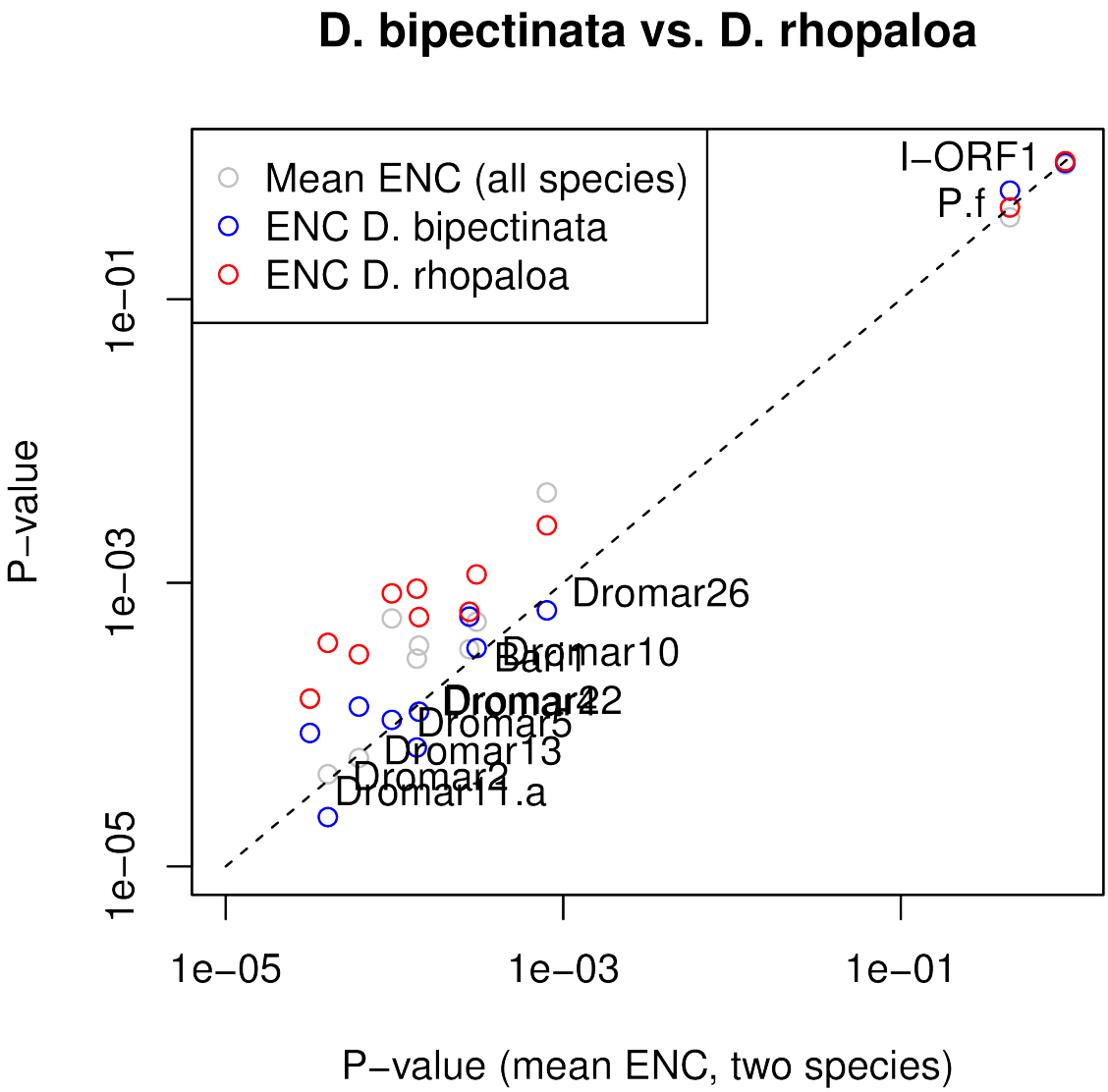


Figure 2: Example of the influence of the choice of ENC on the p-value of horizontal transfers. The p-value obtained for the average ENC between species pairs stands on the X axis. The p-value obtained from other ENC measures are indicated on the Y axis (gray: mean ENC over all species, blue and red: ENC from a single species). The dotted line represents the 1:1 line (no effect).

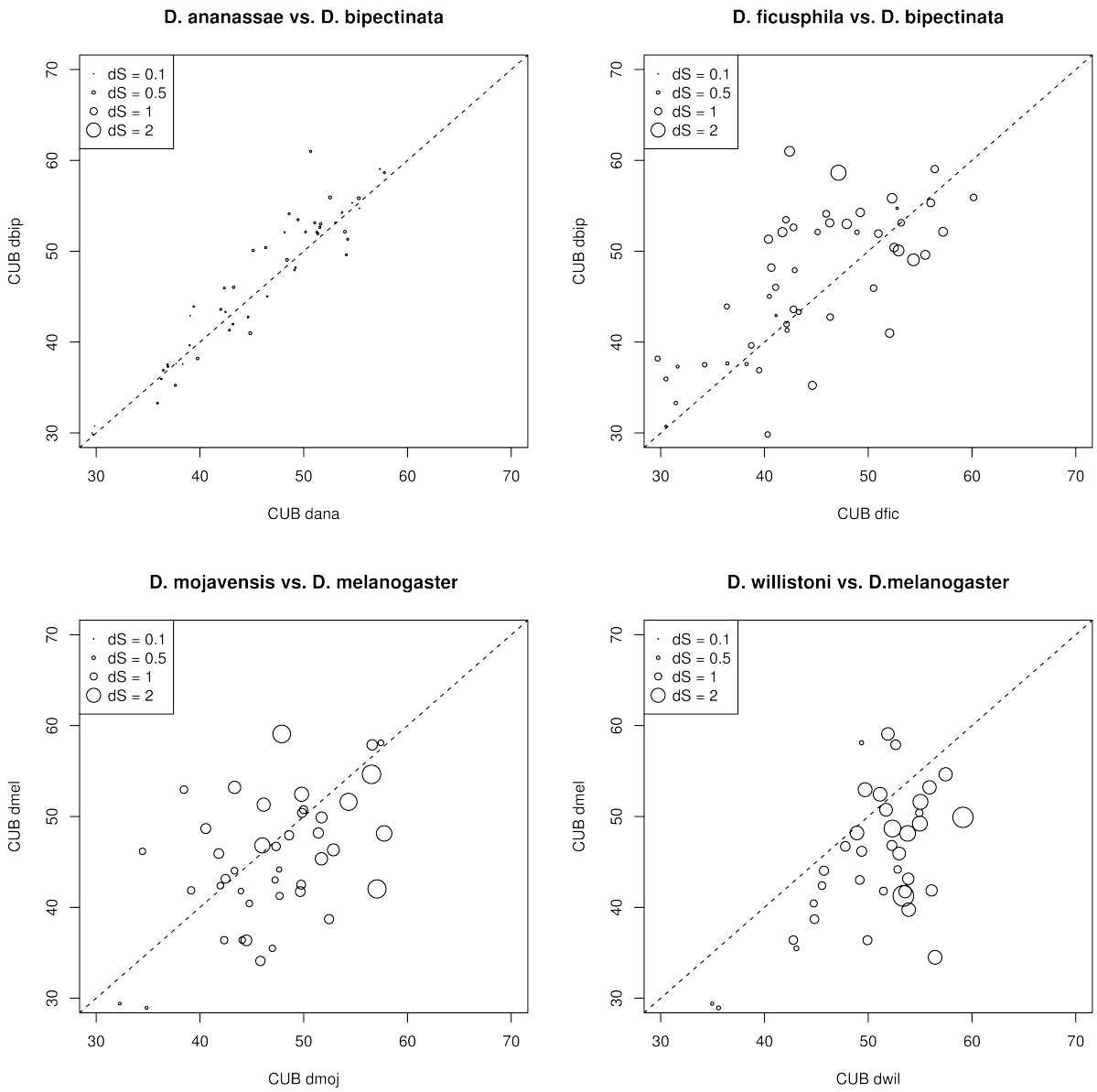


Figure 3: Relation between Codon Usage Bias (ENC) in pairs of species, and gene divergence. The dS is proportional to the symbol size. The dotted line stands for the 1:1 expectation.

Supplementary material S4

Testing Gaussian Residuals

The VHICA method relies on a series of statistical assumptions, among which one of the most important is the Gaussian distribution of the residuals of the regression between codon usage bias (measured as the Effective Number of Codons, ENC) and the synonymous divergence dS. Indeed, P-values of the statistical test (H_0: vertical transmission, H_1: horizontal transfer) are calculated assuming that genes are vertically transferred, and that residuals are normally distributed.

Further investigation showed that the normalized residuals are very close to a Gaussian distribution, and especially that this distribution is virtually symmetric (no skew), although slightly leptokurtik (overpeaked). Note that the cumulative distributions overlap almost perfectly around - 2 standard deviations, where the traditional p-value threshold lies.

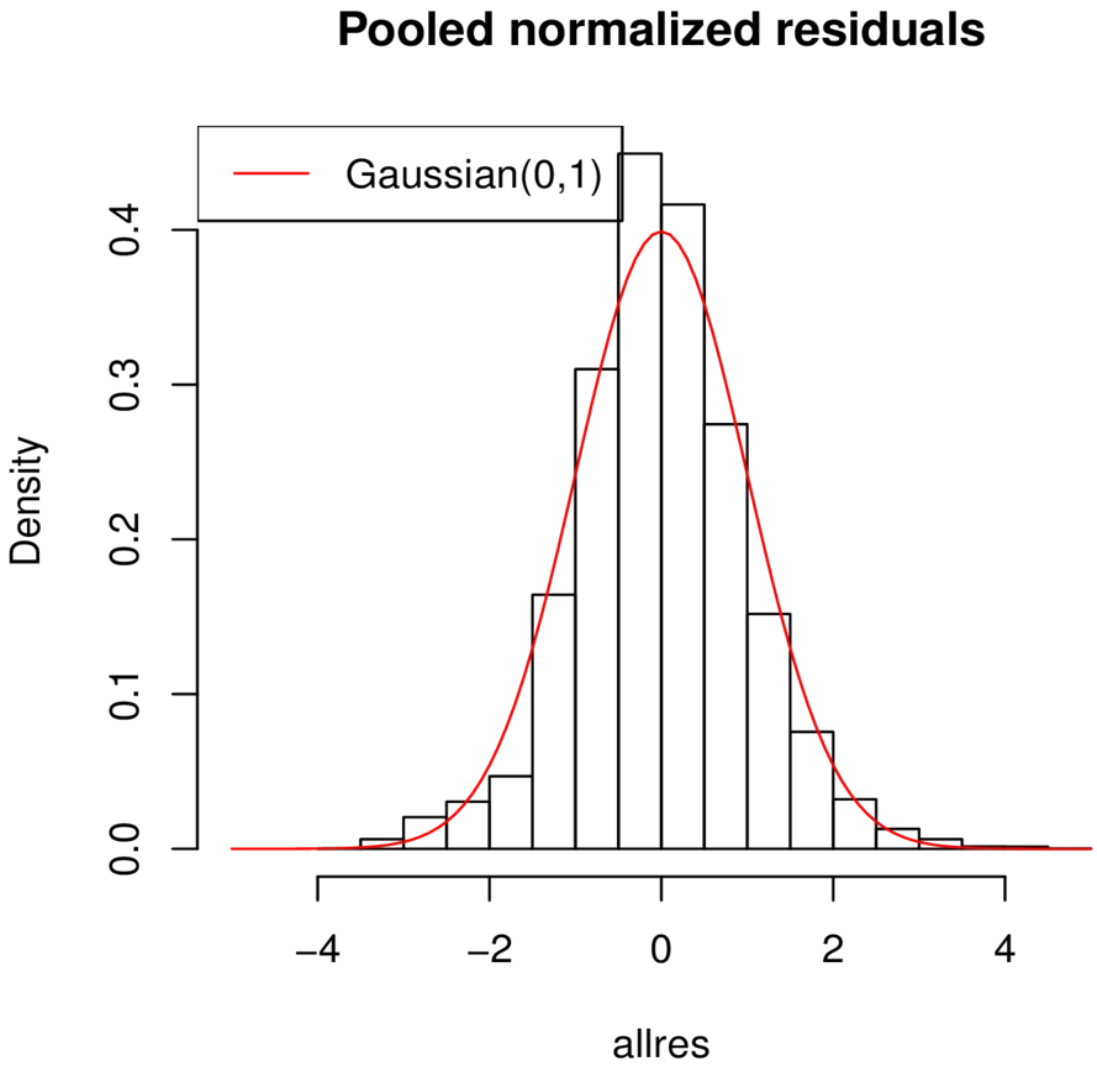


Figure 1: Distribution of normalized residuals (residuals / residual std.dev) pooled across all species comparisons. The red line corresponds to the Gaussian distribution.

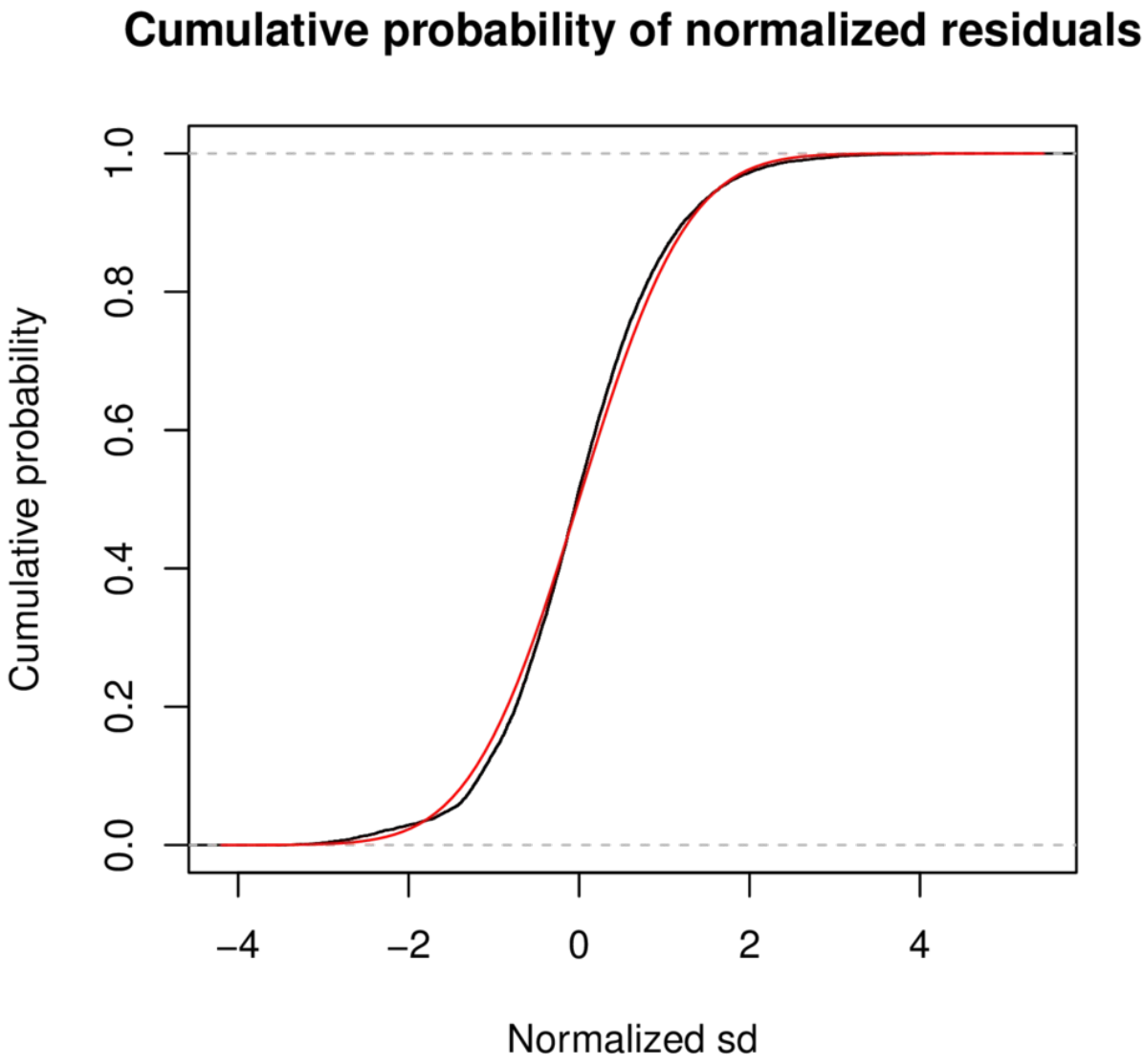


Figure 2: Cumulative distribution of pooled normalized residuals: from the data set in black, Gaussian distribution in red.

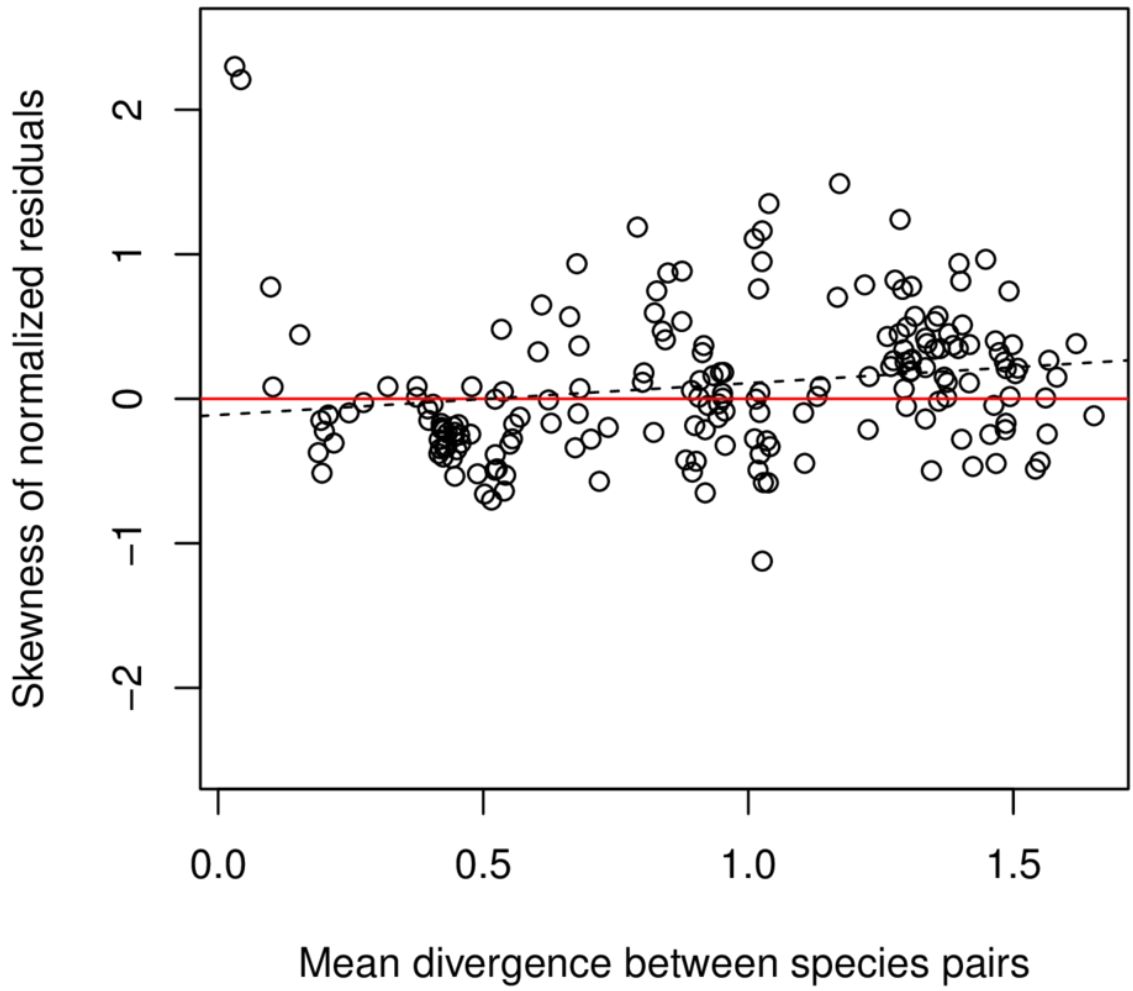


Figure 3: The asymmetry of residual distribution is very close to 0, and is almost not affected by the species divergence. The trend line corresponds to a linear regression, the red line represents the Gaussian expectation.

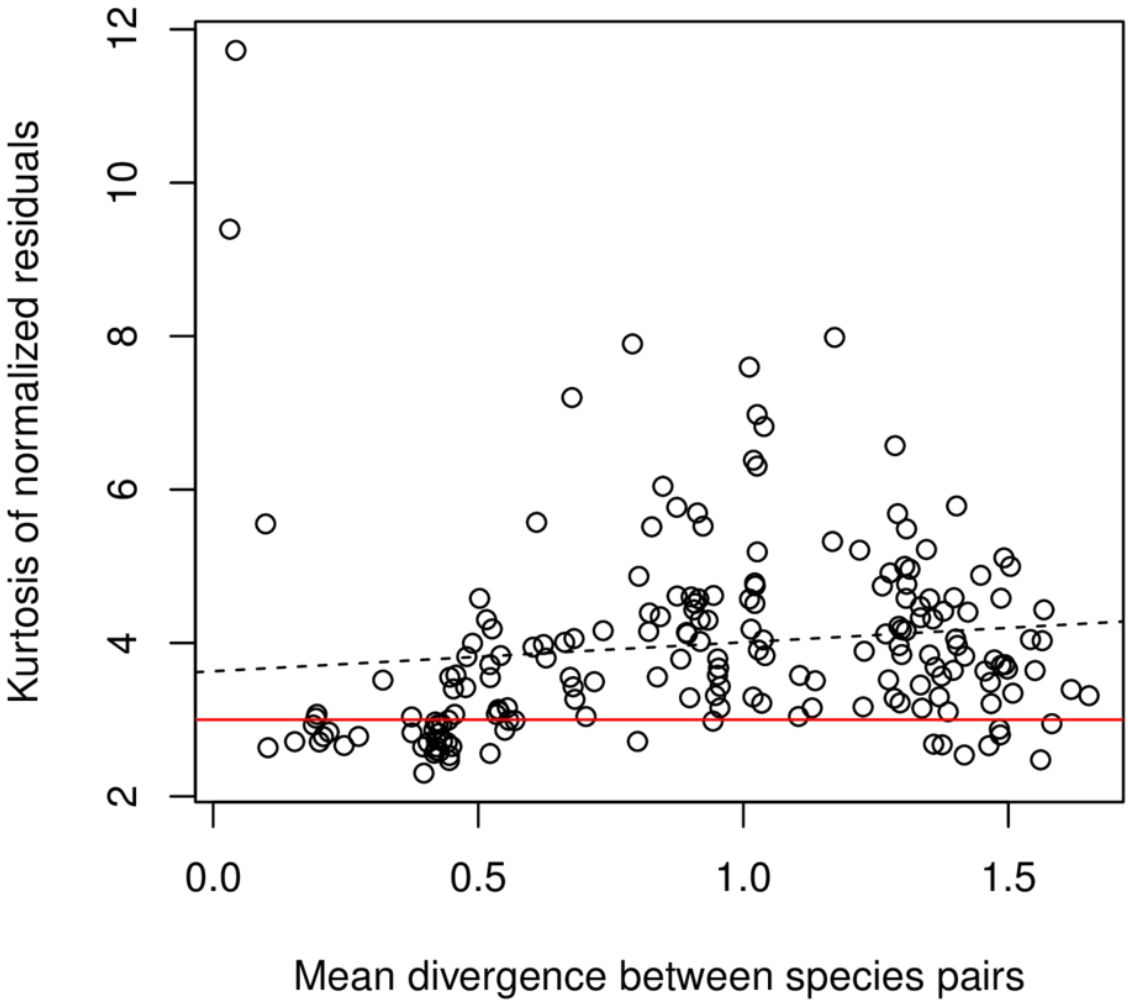


Figure 4: Kurtosis of the normalized residuals vs. divergence between species. The trend line is a linear regression, the red line represents the Gaussian expectation (kurtosis = 3).

Supplementary material S5

Influence of the number of genes

The VHICA method relies on the computation of the deviation of TE sequences from a reference ENC vs dS trend. This reference trend is determined from a set of genes, expected to represent the average relationship between codon bias and divergence during evolution.

The statistical procedure assumes that the reference trend is known perfectly. This is an important approximation, as the trend is actually computed from a finite set of genes (50 in our data set). In order to assess the influence of sampling effects on the VHICA results, we ran two series of resampling tests.

Figure S4.1 shows the distribution of horizontal-transfer P-values with 100 random sets of 50 genes (resampled with replacement in a set of 100 genes), focusing on two species sharing many TE sequences (*D. ananassae* and *D. bipectinata*). Although p-values are not exactly equal across replicates, the interquartile range remains small enough to ensure consistent results for the majority of cases: among 26 TE sublineages, 5 were always (or almost always) considered as significant HTs at the threshold of 0.05, and 12 were always considered as vertically-transfered.

Figure S4.2 show the frequency at which HTs are detected for three of these TEs as a function of the number of genes (*Dromar22* is a certain horizontal transfer, *Dromar10* is a certain vertical transfer, and *Dromar24* is a doubtful case). Interestingly, the only impact of increasing the number of genes is to discard false positives (*Dromar10* is sometimes tagged as a horizontal transfer when less than 20 genes are considered). In particular, whether or not *Dromar24* is horizontally transfered is not fixed when increasing the number of genes.

In sum, it seems reasonable to include more than 30 genes to compute the regression, as fewer genes increases the rate of false positives. However, there is no obvious benefits in increasing the number of genes further (no gain in power).

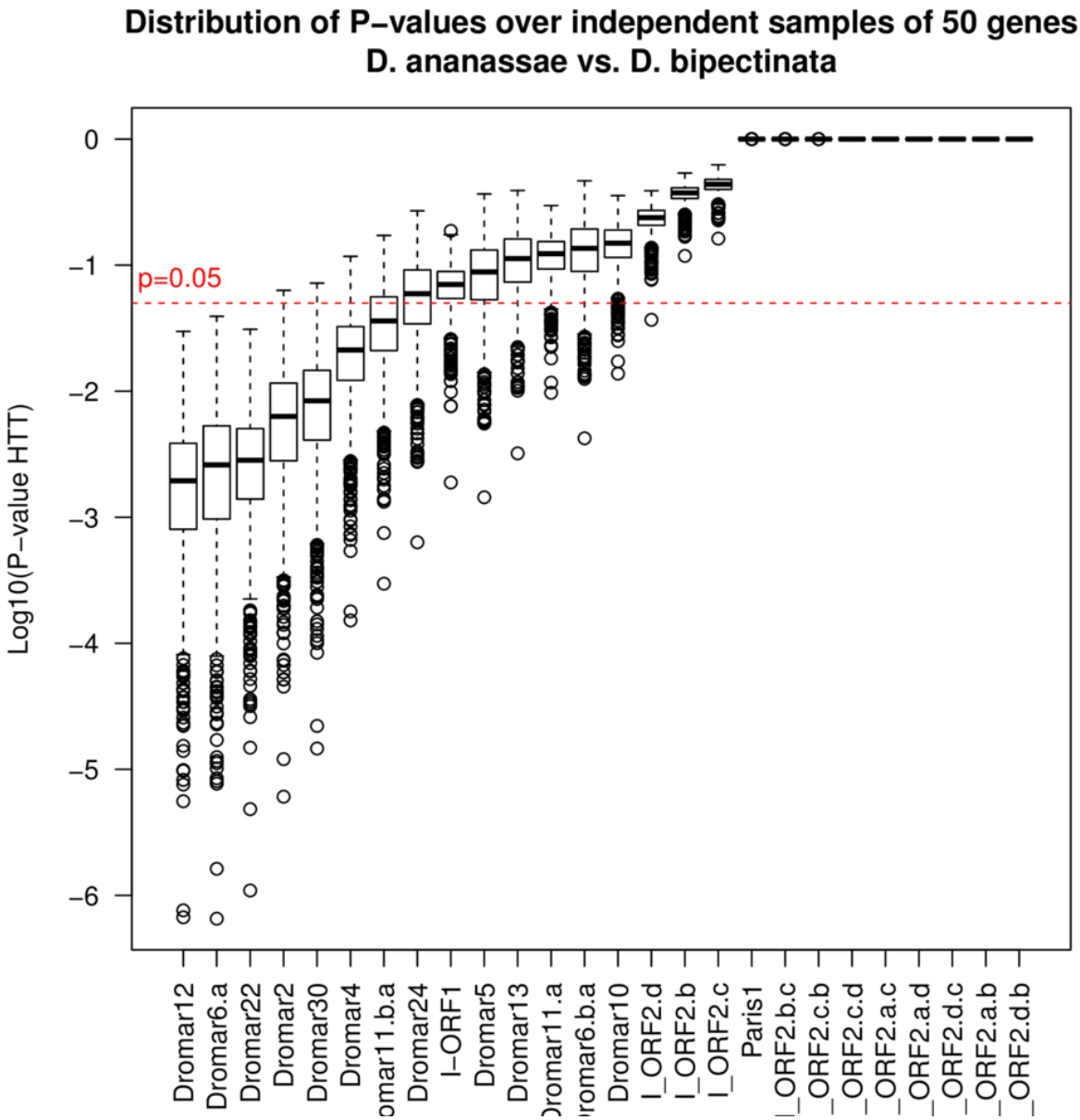


Figure 1: Distribution of p-values (H0: vertical transmission, H1: horizontal transfer) under random samples of 50 genes.

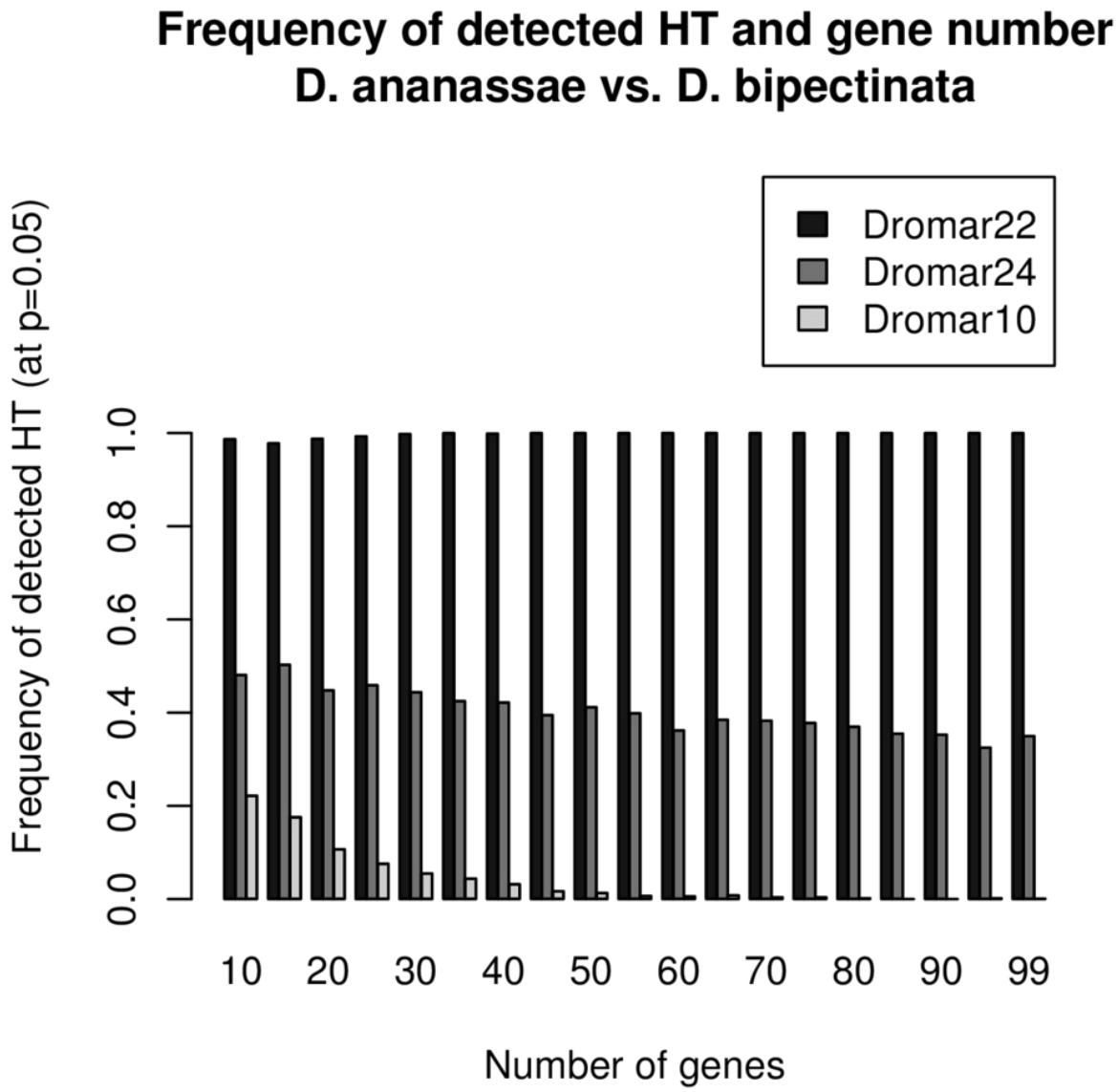


Figure 2: Frequency of HT detection vs the number of genes.

Supplementary Material S6.

Choice of the representative sequence.

Compared to unique genes, the choice of a TE sequence among several sequences from the same genome can be problematic. The existence of different clades (sublineages), as revealed from the phylogenetic analysis, supposes that one representative of each sublineage should be tested.

The simplest method consists in choosing the most complete sequence, that contains the less deletions and the less insertions compared to the consensus sequence derived from the whole dataset (all copies in all species). But for some TE lineages, most sequences are complete and have equal size. Furthermore, the comparison to the consensus sequence, which often reflects a more ancestral state, may bias the choice toward slowly evolving sequences.

This method may then be unsuitable, and we ran different tests in order to evaluate the influence of the TE representative sequence choice, on various TEs analyzed in this work.

Three different tests have been performed:

1- Selection of sequences on the basis of the dS or the ENC

For this, we selected within each species (and within each sublineage if needed) the sequence(s) exhibiting the dS, or the ENC closest to the average dS or ENC. To identify average dS and ENC, vhica was run considering all sequences in the TE alignment. The average ENC for one species is easy to compute, since it depends only on the sequence. However, sequences are associated to different dS values, depending on the second sequence used in the comparison. Then we computed the average dS for each species pair, extracted the sequence exhibiting the closest dS to these averages, and then chose for each species the sequence most frequently selected in the different comparisons.

For all elements tested (Figure S6.1), the resulting p-value matrices showed not significant difference, suggesting that choosing one random sequence among the most complete is a fair compromise.

2- Results obtained for randomly chosen sequences.

In this second test, we focused on some typical species comparisons, selected for displaying HTT, or VTT. Hence, three or four pairwise comparisons were analyzed for P, dromar8 and I-ORF1 (Figure S6.2). Sequences were randomly chosen within each species (and within each sublineage, for P element in *D. willistoni*), and 100 replicates were done for each species pair. Although different sequences map differently on the pairwise map, it is obvious that most sequence pairs give results similar to the one obtained when sequence are selected manually.

- For the strongly statistically-supported HTT cases (dwill x dmel for the sublineage a of the P element; dfic x dgri for *dromar8*), all the comparisons resulted in strong HTT signal, the most variable parameter being obviously the ENC.
- For HTT less strongly supported (dere x dfic and dana x dfic for dromar8), only a few sequence comparisons failed to detect the HTT.
- For doubtful HTT (p-value close to the cut-off for significance, such as dana comparison with dmel and dbip, for I-ORF1), all comparisons remained doubtful.

- For comparisons which did not exhibit HTT signal, HTT signal was never detected, whatever the sequences (different sublineages comparison for P elements, dbia x dfic for dromar8 and deug x dmel for I-ORF1)

Hence, it can be concluded that choosing totally randomly a sequence can prevent a HTT detection, only when the HTT is not strongly supported. The risk of false positives or false negatives is then not increased as soon as the departure from genes is far from the cut-off. Note that in the case of the deug x dana comparison of I-ORF1, all replicates give significant HTT, while the p-value is never high.

3- Influence of the sequence length

Finally, we tested the impact of the sequence length on the results by running vhica on artificially truncated sequences. According to Wright (1990), sequences with less than 60 codons should not be analyzed for ENC. Since the ENC is calculated on the whole sequence, and the dS is calculated only on the codons common to the two sequences, the dS comparison can be done on sequences sharing less than 60 codons.

For this test, artificial sequences were created by drawing codons from the real sequences, but contained a fixed number of common codons chosen randomly among comparable codons, with a replacement option. We chose different elements in different species that exhibited either no HTT signal, a strong HTT signal or a doubtful signal. For each pair, 100 replicates have been done with 200 or less codons. Results are shown in fig. S6.3.

For each test, the dS and ENC were first averaged over the 100 replicates with various numbers of codons (60, 100, 150, 200) (Figure S6.3.A). In all cases, we could notice that the averaged dS is poorly affected by the codon number, although the variance increased with the reduction in codon number. On the opposite, the ENC was highly sensitive to the codon number, and found to be drastically lower for low codon numbers. Hence, a small number of codons give rise to biased ENC. It is then very important to check that enough codons are available for ENC calculation.

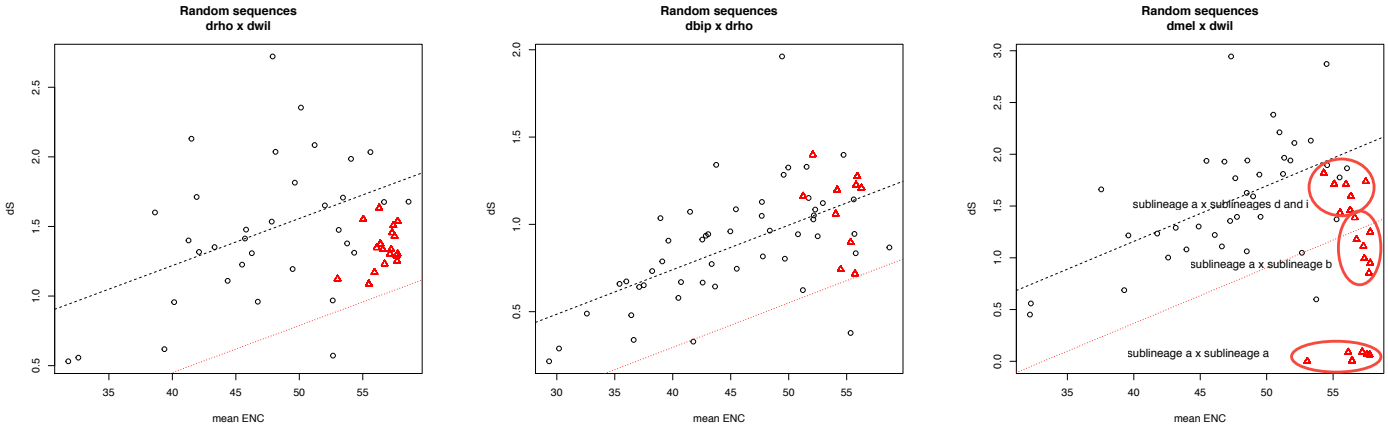
Such biased ENC makes vhica less sensitive, since when the ENC is low (biased), a lower dS is needed for the comparison to significantly depart from the gene values. It is then expected that small codon number will generate mainly false negatives. This is what was observed after analysis of the sampled sequences.

For sequence pairs identified as vertically transmitted (not HTT signal), all replicates gave p-value above the cut-off of 0.05 (figure S6.3.B). For sequences identified as clear HTT, a small fraction of replicates were above the cut-off, when 100 codons were used (figure S6.3.C). Finally, for less supported HTT (HTT signal fading when Bonferroni correction is used), the HTT signal is lost in most replicates with 100 codons (figure S6.3.D), which is expected.

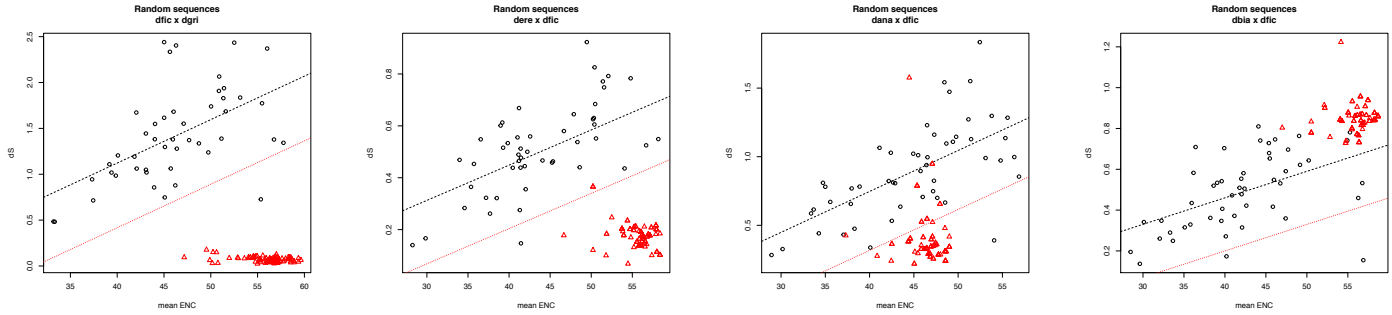
Hence, the use of too few codons never increases the rate of false positives, but likely increases the rate of false negatives, when the HTT signal is weak. It seems then important to use as much sequences as possible in the analysis in order to keep enough sensitivity. Note however that vhica performs ENC calculation on the whole sequences, whatever the number of codons used in the dS analysis. Since the dS value are quite stable, the most important in the length of the sequences and not the length of the shared codons. Then comparison of sequences, long enough, but sharing few codons, is still possible.

Figure S6.1 : vhica results obtained with arbitrary or selected sequences for P element (A), Dromar8 (B) and the I-ORF1 (C). For these tests, the Bonferroni correction was systematically applied.

A- *P* element



B- *dromar8*



C- I Factor - ORF1

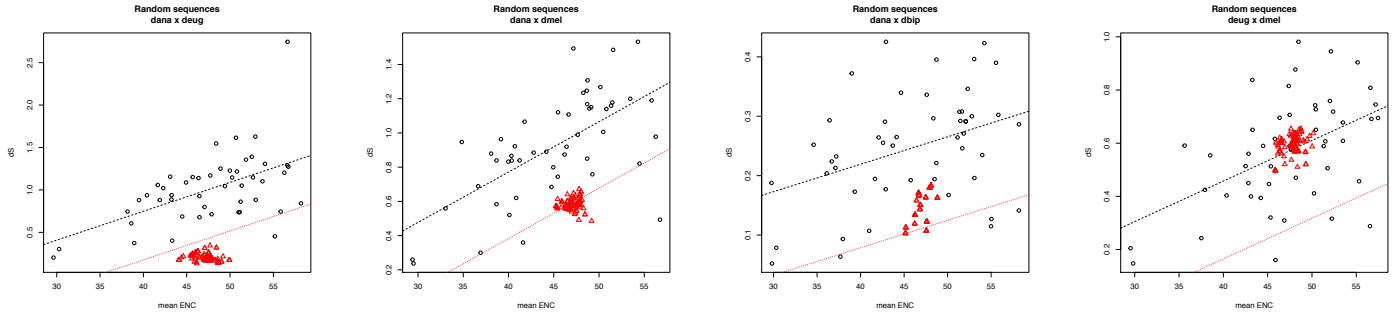
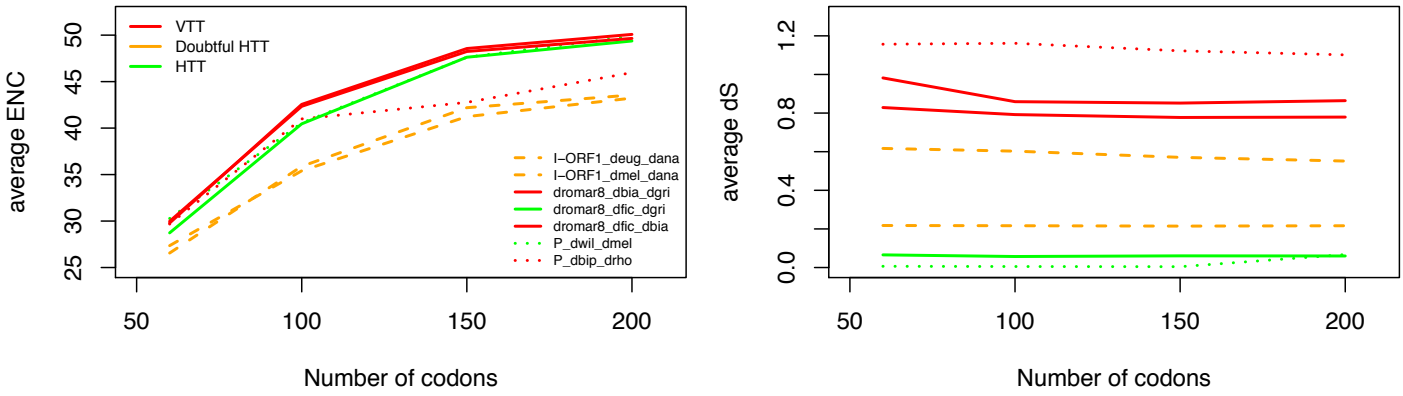


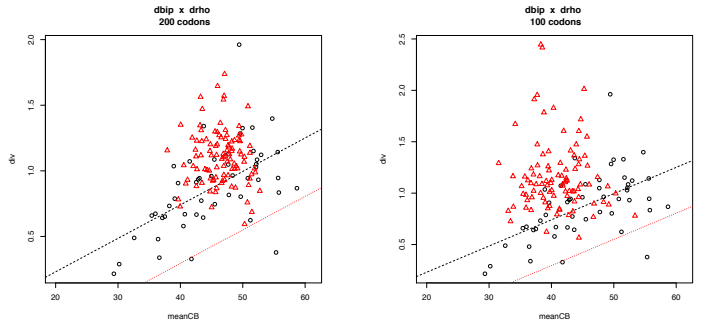
Figure S6.2 : vhica results obtained with random sequences picked up from the alignment of all copies, for P (A), Dromar8 (B), I-ORF1 (C). TE values are displayed as red triangles, each triangle being one of the 100 replicates.

A- Influence of the number of codons on the ENC and the dS

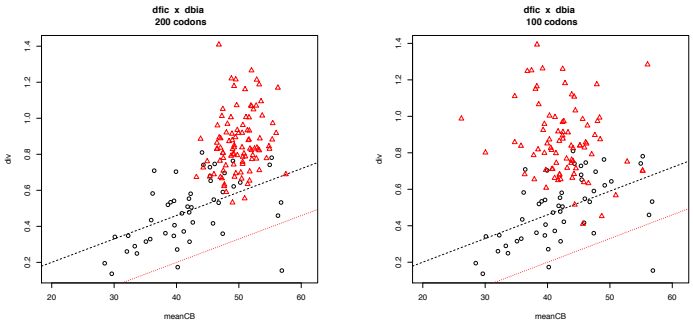


B- Examples of VTT

P element dbip x drho
(246 codons)

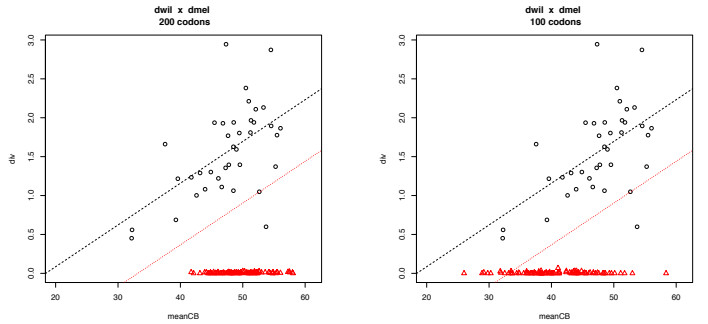


dromar8 dfic x dbia
(343 codons)

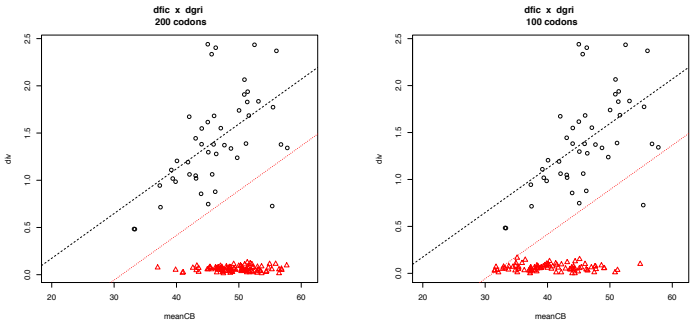


C- Examples of HTT

P element dwil x dmel
(750 codons)

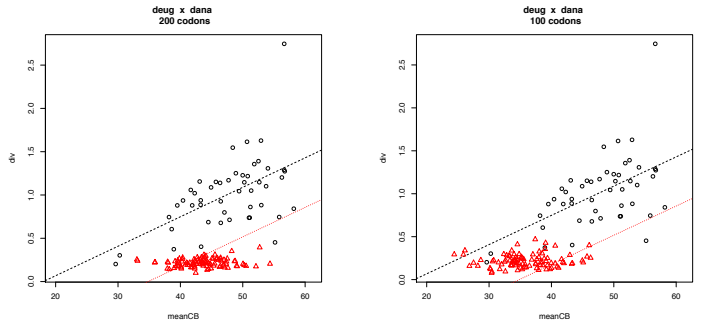


dromar8 dfic x dgri
(352 codons)



D- Examples of less supported HTT

I-ORF1 deug x dana
(414 codons)



Chouto-POL dmel x dsim
(773 codons)

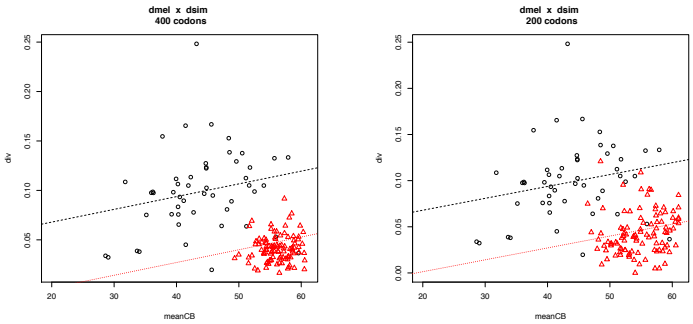


Figure S6.3 : Influence of the number of codons. A) en ENC and dS. B) to D). vhica results for various TEs for which 100, 200 or 400 codons were resampled (100 replicates)