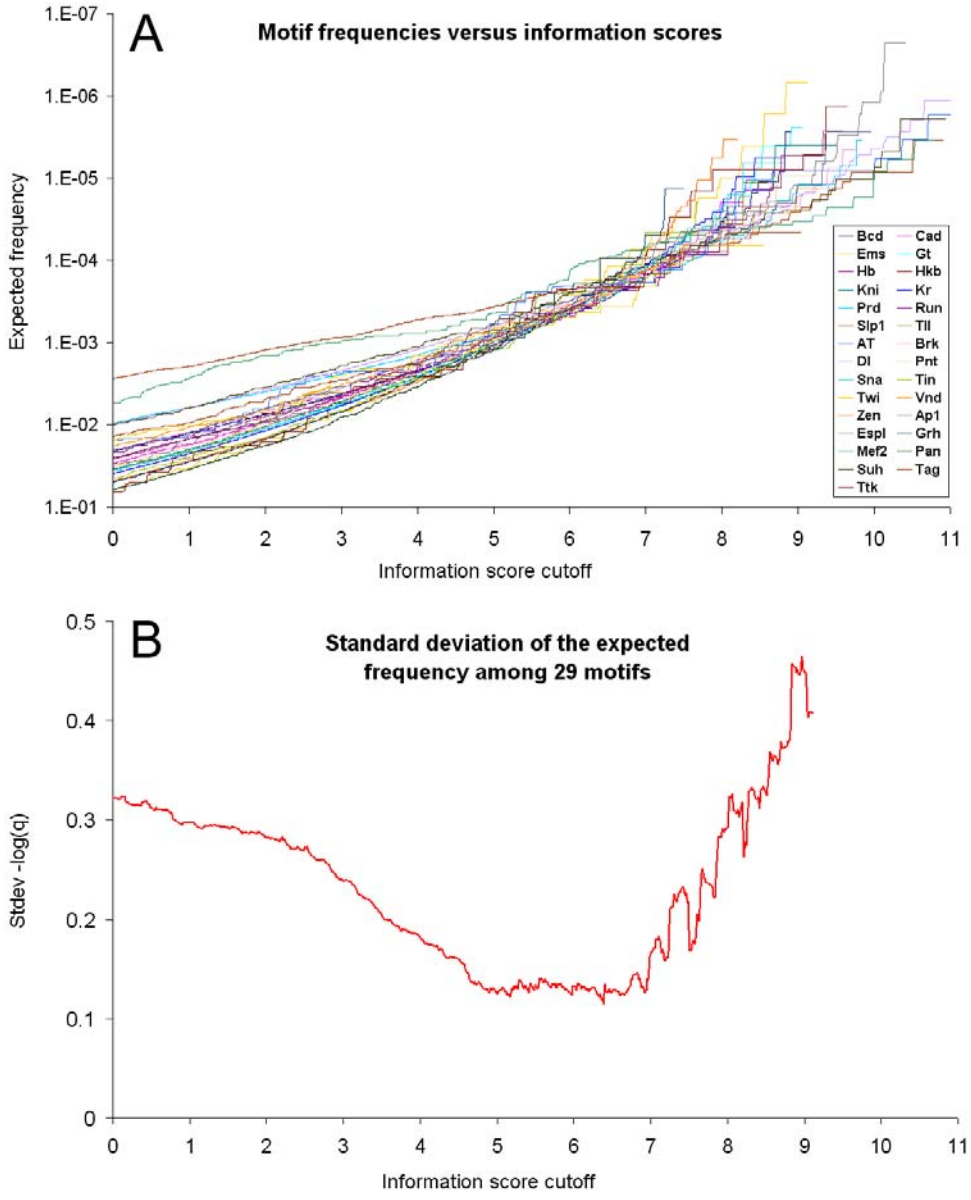
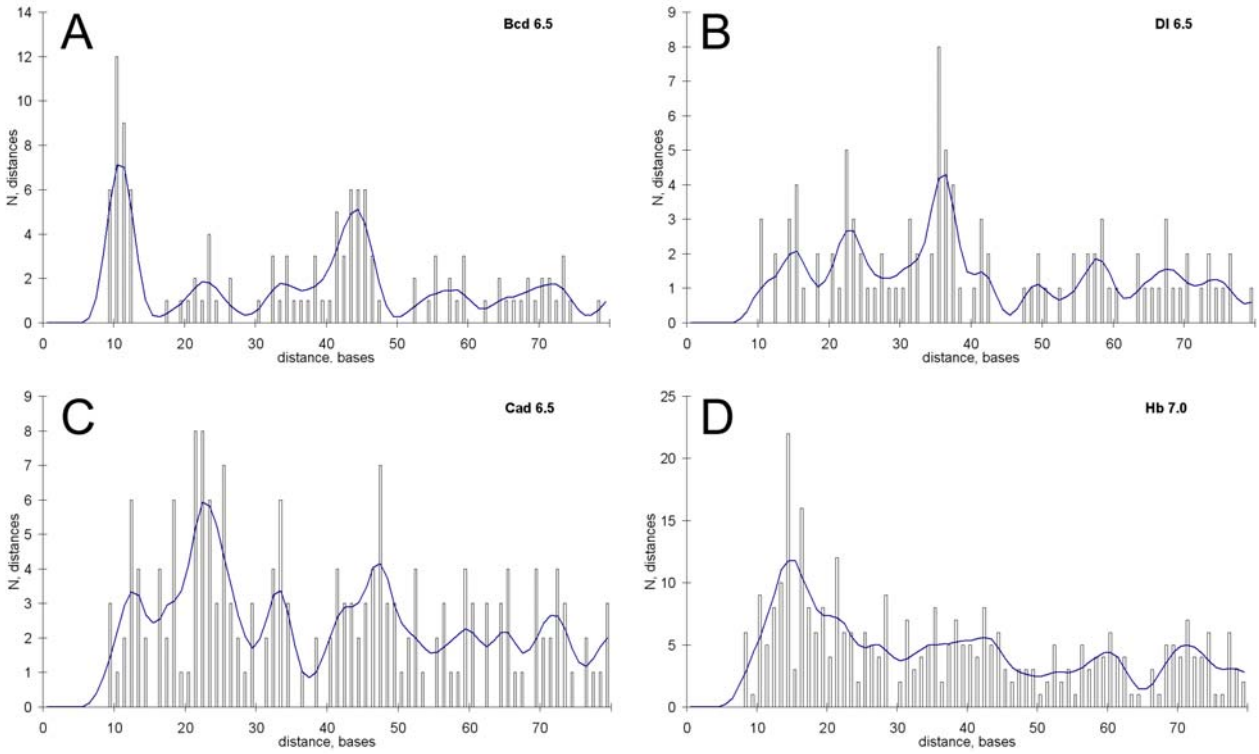
Organization of developmental enhancers in the Drosophila embryo

D. Papatsenko, Y. Goltsev and M. Levine

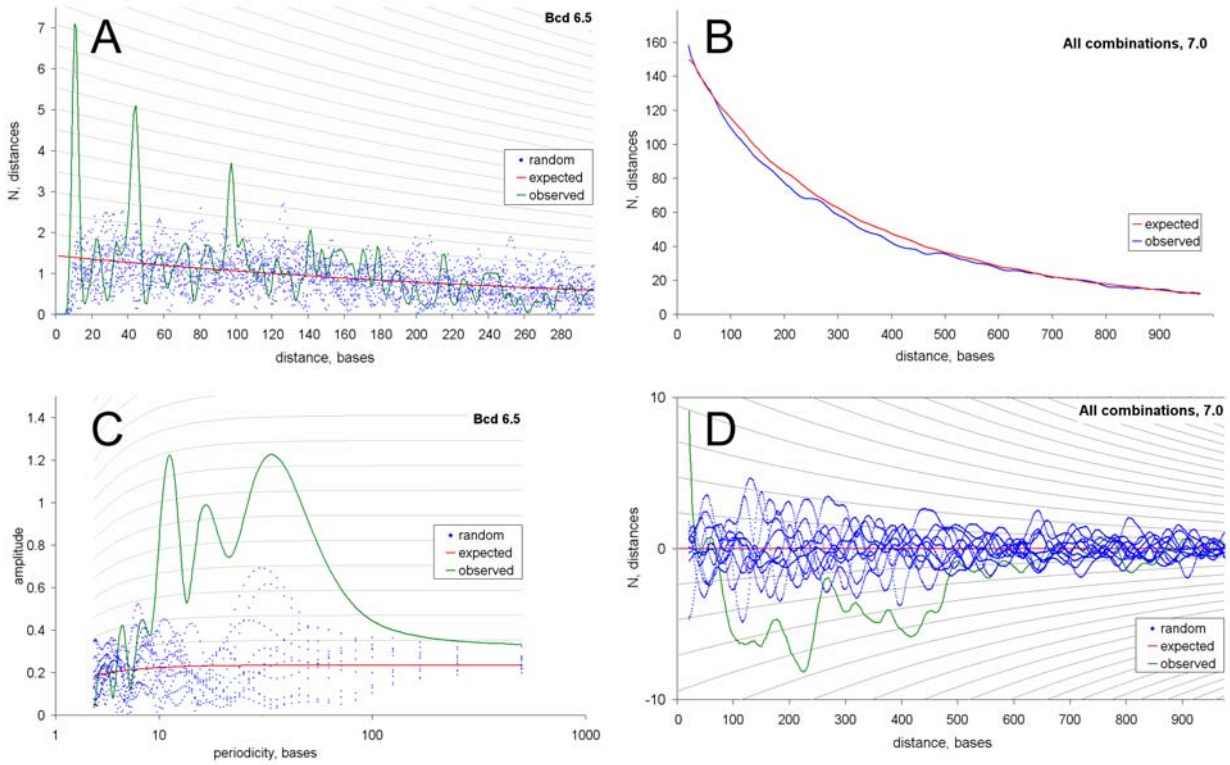
Supplemental data



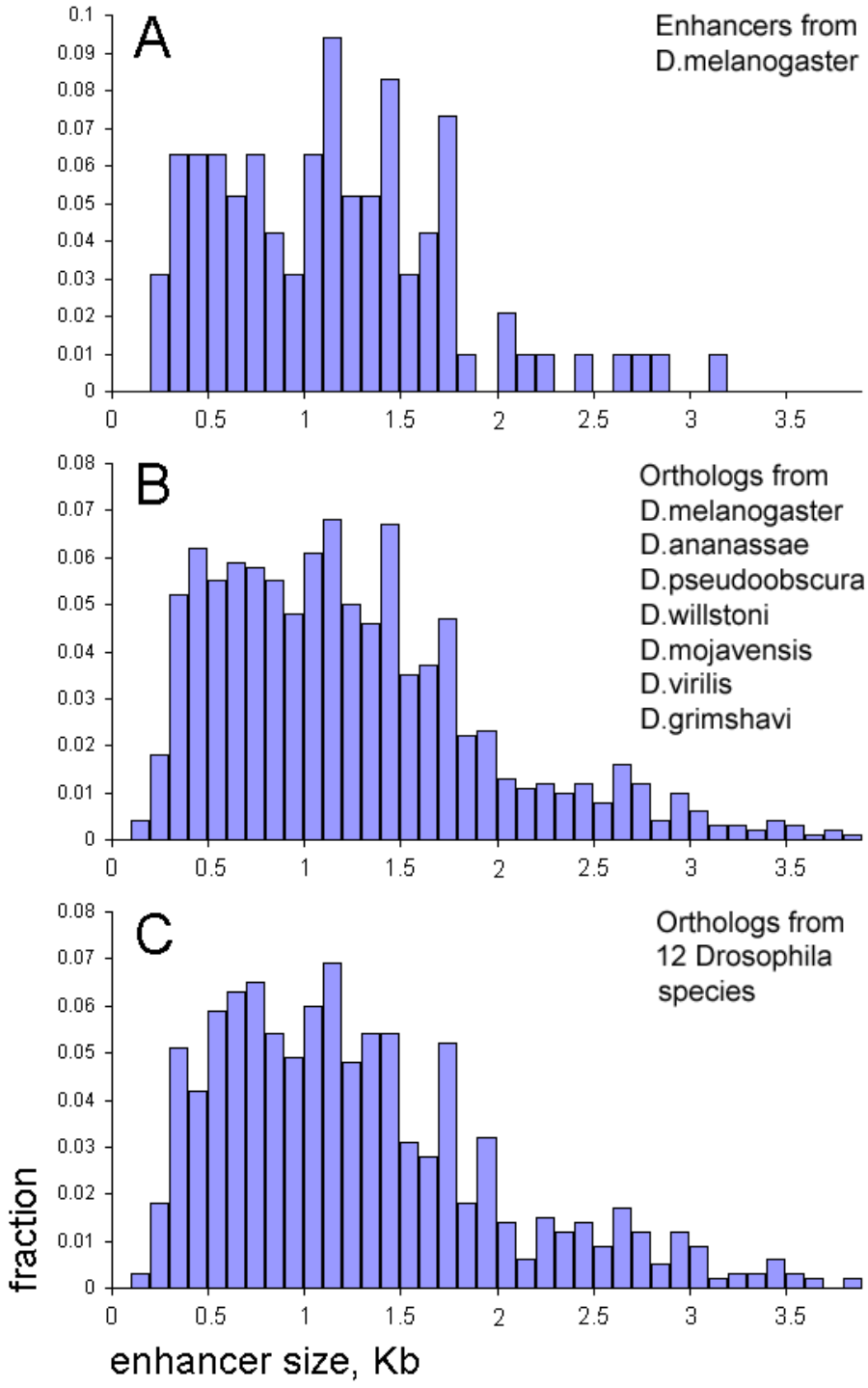
Supplemental Figure S1. Selection of informational cutoff for motif matches
(*A*) Dependence of the expected match frequency from informational cutoff for 29 binding motifs used in this study. (*B*) Standard deviation of –log frequency, calculated for the 29 motifs. For the majority of motifs, frequency is similar in the range 5-7 bits. In this range, the same match cutoff will produce similar number of matches for nearly any binding motif.



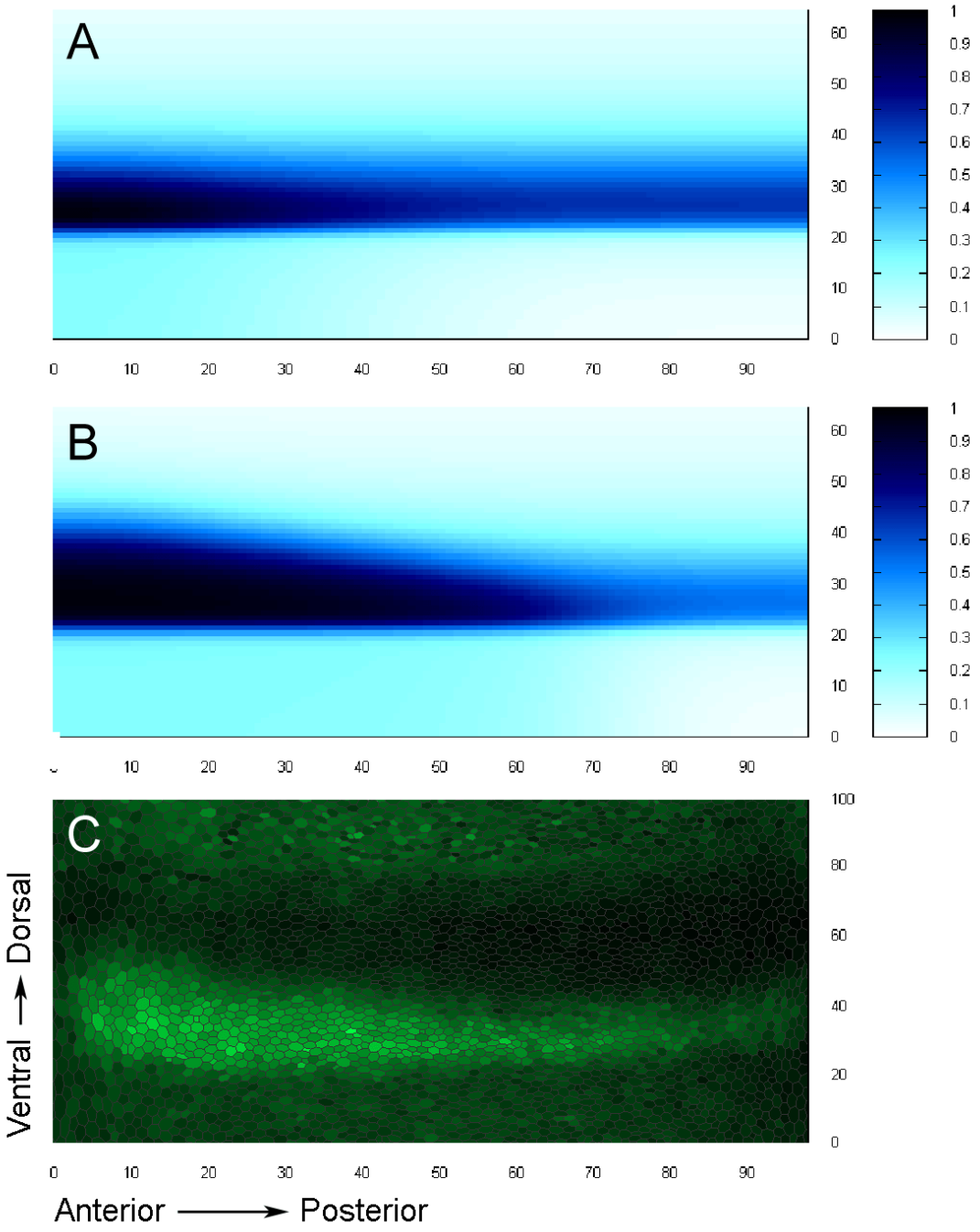
Supplemental Figure S2. Short-range distance histograms
Short-range distance histograms for homotypic motif combinations: (*A*) Bicoid-Bicoid, (*B*) Dorsal-Dorsal, (*C*) Caudal-Caudal, (*D*) Hunchback-Hunchback. Three out of four transcription factors (*A*-*C*) demonstrate presence of periodic signal ~11 bp in the distribution of binding sites.



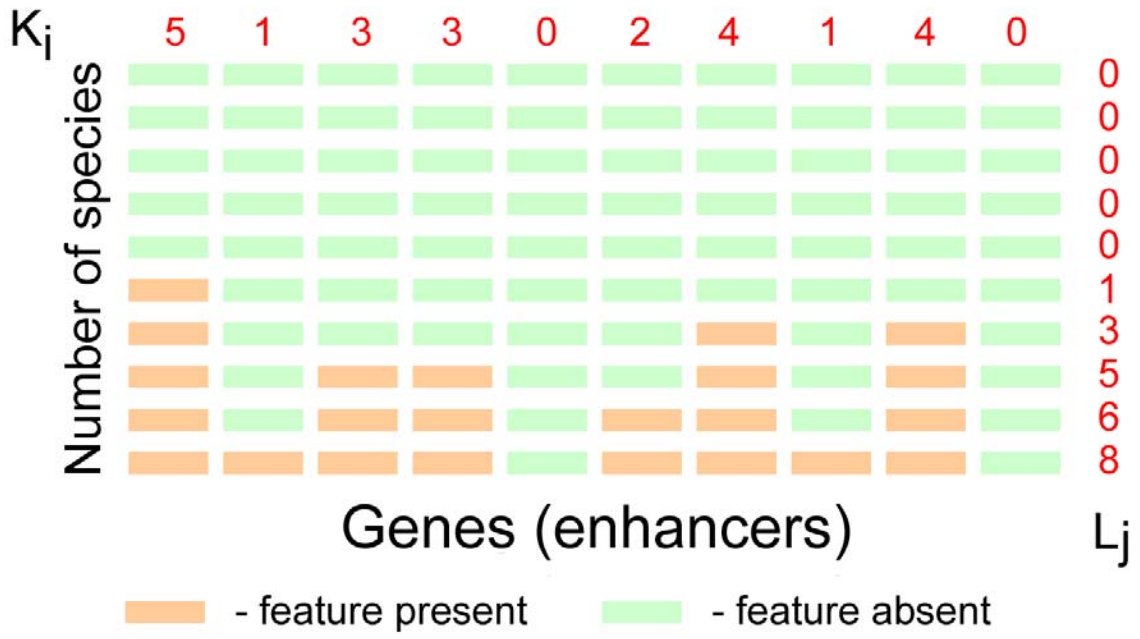
Supplemental Figure S3. Statistical evaluation of distance histograms
(*A*) Distance histogram for Bicoid is shown in comparison with 10 histograms, computed using the same set of Bicoid sites, positions of which were randomized in each enhancer. (*B*) Comparison of the expected and observed long-range distance histograms for all motif combinations. (*C*) Fourier spectrum for Bicoid distance histogram shown in panel (*A*) (0-80 bases) in comparison with Fourier spectra obtained after randomization of positions of the Bicoid matches in enhancers. (*D*) Differential long-range distance histogram for all motif combinations in comparison with differential histograms obtained by the match position randomization. Gridlines on panels (*A*, *B*, *D*) are given in standard deviations, computed based on the corresponding randomization tests.



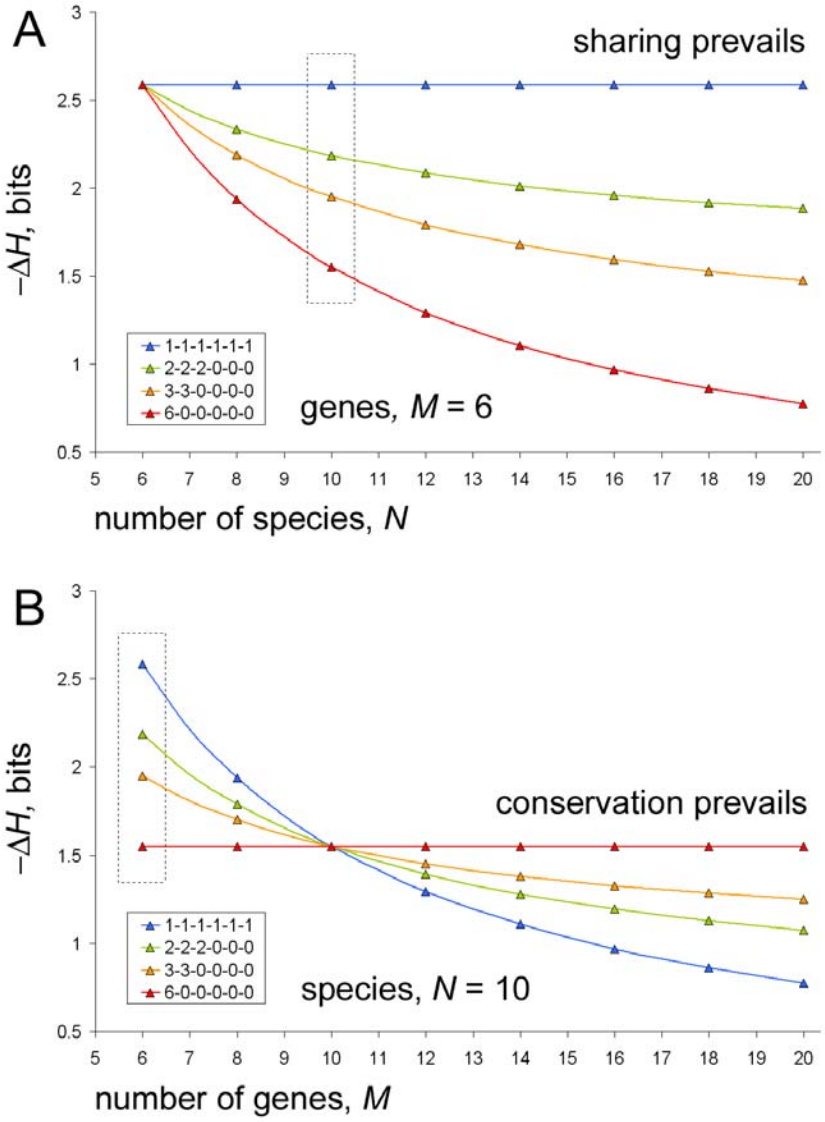
Supplemental Figure S4. Distribution of enhancers by size
(*A*) The original enhancers, identified in D.melanogaster, (*B*) Orthologs from 7 most distant species, used in this study (*C*) orthologs from 12 Drosophila species.



Supplemental Figure S5. Model for Anterior-posterior modulation of Brinker
The model was constructed using site-occupancy models described before. Inputs include Bicoid, Dorsal and Snail gradients. Assumptions include the following: 1. Dorsal is required for expression of Brinker, 2 Bicoid is not required, but it helps Dorsal binding, 3 Binding of Snail abolishes expression. (*A*, *B*) Increasing Bicoid-Dorsal cooperativity by order of magnitude extends Brinker expression to posterior. (*C*) Orthographic projection of Brinker expression (from BDTNP expression atlas).



Supplemental Figure S6. Calculating entropy scores
Motif combination (feature) is present only in some species and only in some enhancers (groups). Entropy of conservation is computed from the number of orthologs K_i, containing the feature in each enhancer (group of orthologs). Entropy of sharing is computed from the maximal possible penetration L_j of the feature across different groups. As long as all species are equivalent (more or less evolutionary equidistant), it is not critical in which exactly species the motif combination (feature) is present.



Supplemental Figure S7. Score dependence from the number of genes and species
Entropy scores are shown for four different features with different levels of conservation and sharing: the first feature (in blue) is shared by 6 enhancers, but it is not conserved: one instance is present in each enhancer. The second feature (in green) is shared only by 3 enhancers, and conserved in only 2 orthologs of the enhancers. The third feature (in yellow) is shared by two enhancers and conserved in 3 orthologs. The last feature (in red) is present only in one enhancer, but conserved in all 6 species. Depending on the total number of species and enhancers under consideration, either conservation or sharing scores prevail. (*A*) With increasing of the number of species in the test, the conservation level 6 (red feature) becomes insignificant and the score of the most shared, even not conserved features (blue) prevails. (*B*) With increasing of the number of genes in the test, sharing level 6 (blue feature) becomes insignificant and scores of more conserved features prevail.

Table S1. Nucleosome formation potential in enhancers and non-functional sequences

The top table contains information for 14 functional enhancers, the bottom table contains information for 9 non-functional sequences, both types of sequences have been tested in vivo. Nucleosome formation potential has been measured for each enhancer sequence using Recon program. Difference between distributions of the average nucleosome formation potential in the two groups is significant (T-test p=0.012). The last column shows correlation between the nucleosome formation potential and enhancer borders (function: enhancer sequence = 1, flank sequence = 0). 11 out of 14 functional enhancers have negative correlation between the nucleosome formation potential and enhancer borders. Most of the false-positive sequences (6 out of 9) have positive correlation.

Enhancer Name	Function *in vivo*	Fragment Size (bp)	Av. Nucl. potential	Frag.size+ flanks	Correlation (*r*)
eve 3+7	enc	511	0.19	1548	-0.36
eve 2	enc	730	0.24	1767	-0.03
eve 4+6	enc	602	0.60	1639	0.34
eve 1,5	enc	1798	0.37	1806	-0.05
brk	enc	501	0.48	1539	-0.12
brk-s	enc	1022	0.54	2060	-0.04
CG12177	enc	336	-0.06	1374	-0.06
rho	enc	302	0.36	1340	-0.20
sim	enc	634	0.30	1752	-0.34
sog	enc	395	-0.04	1413	-0.23
sog-s	enc	887	0.33	1925	-0.38
ths	enc	511	0.69	1549	0.21
vn	enc	500	0.35	1538	-0.25
vnd	enc	746	0.35	1784	0.00
average		676	**0.34**	1645	-0.11
CG1412	-	455	0.26	1493	-0.35
CG1924	-	396	0.51	1434	0.02
CG5549	-	508	0.58	1546	0.19
eya	-	169	0.48	1207	-0.09
fas-3	-	188	0.63	1226	0.01
Kr-h2	-	337	0.70	1375	0.21
PpD5	-	416	0.69	1533	0.08
Ppv	-	121	0.66	1159	0.07
run	-	423	0.33	1461	-0.34
average		335	**0.54**	1382	-0.02
Genome average			**0.46**		