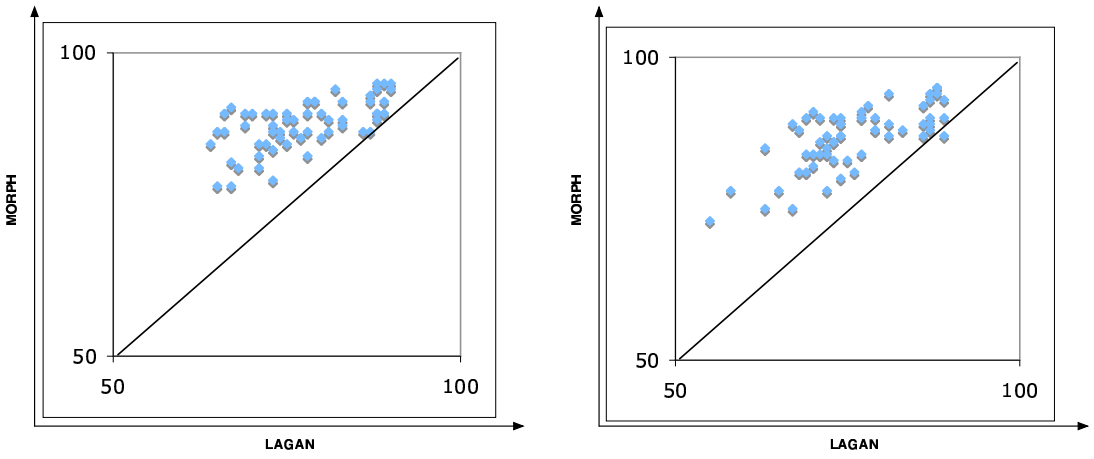

(A) (B)

Figure S1: Alignment sensitivity (A) and specificity (B) of Lagan and Morphalign on experiment sets generated by new simulation method. Diagonal lines represent equal scores.

1 Experiments with Synthetic Data

The results of Figure 1 (main text) were obtained on synthetic data sets where "orthologous CRMs" were generated by sampling from the MORPH probabilistic model. Now, we used a different method to create such "orthologous" pairs of CRMs, so as to ensure that the improvements reported in Figure 1 were not simply due to the fact that the same model was used to generate data sets and predict alignments.

Here, we started with the DAWG program (Cartwright, 2005, *Bioinformatics* 21) designed for simulating non-coding sequence evolution. This program takes any given sequence and "evolves" it under suitable models of nucleotide substitution and indel generation. It does not implement a model of CRM evolution, however. That is, it supports no notion of constrained sites (e.g., TF binding sites). and consequently does not incorporate any binding site loss or gain, which is partly what the MORPH model tries to capture. We therefore implemented our own simulation program that (i) intersperses randomly sampled binding sites with randomly generated background sequence, (ii) evolves binding sites according to an evolutionary model (F81 with PWM as stationary distribution), and (iii) evolves background sequence using DAWG.

More specifically, sequences are created by alternating between binding sites and background or "spacer" regions. The binding sites are sampled from one of seven PWMs (as in Section 2.2), and evolved according to an evolutionary model which is equivalent to Felsenstein 81 with PWM frequencies as the stationary distribution. The spacer regions have length chosen from a geometric distribution with mean γ, and evolved using DAWG. Nucleotides are sampled according to background frequencies, and evolved according to an F81 model again. Indels are generated from a power law distribution with exponent ϵ and a maximum indel size of 20. The spacer region mean γ and the indel distribution parameter ϵ are experimental variables and are set to different combinations of values. Sixty data sets are created in this way, and both Morphalign and Lagan (with default parameters) are run on each data set. The sensitivity and specificity of alignment, for each method, is shown in Figure S1.