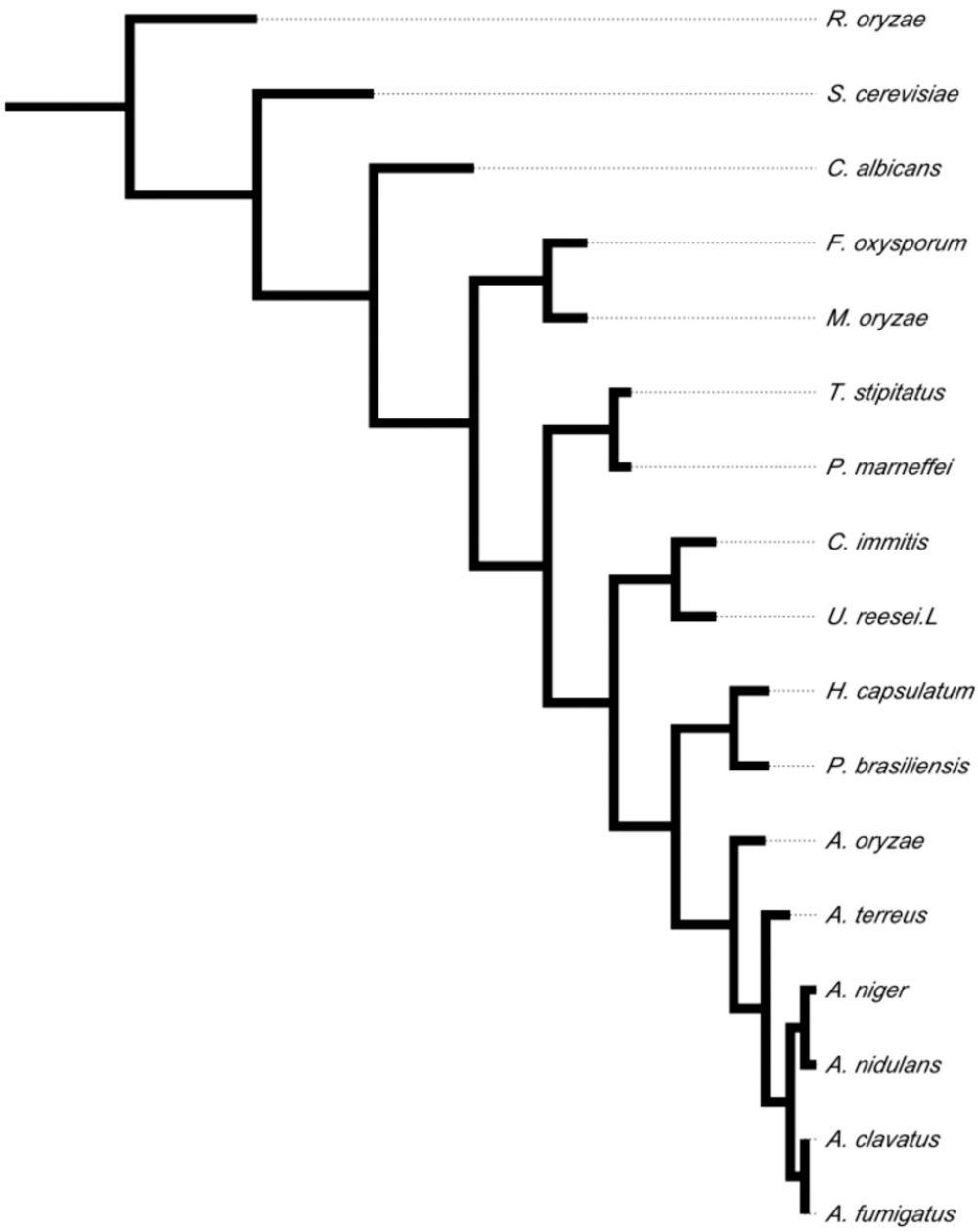
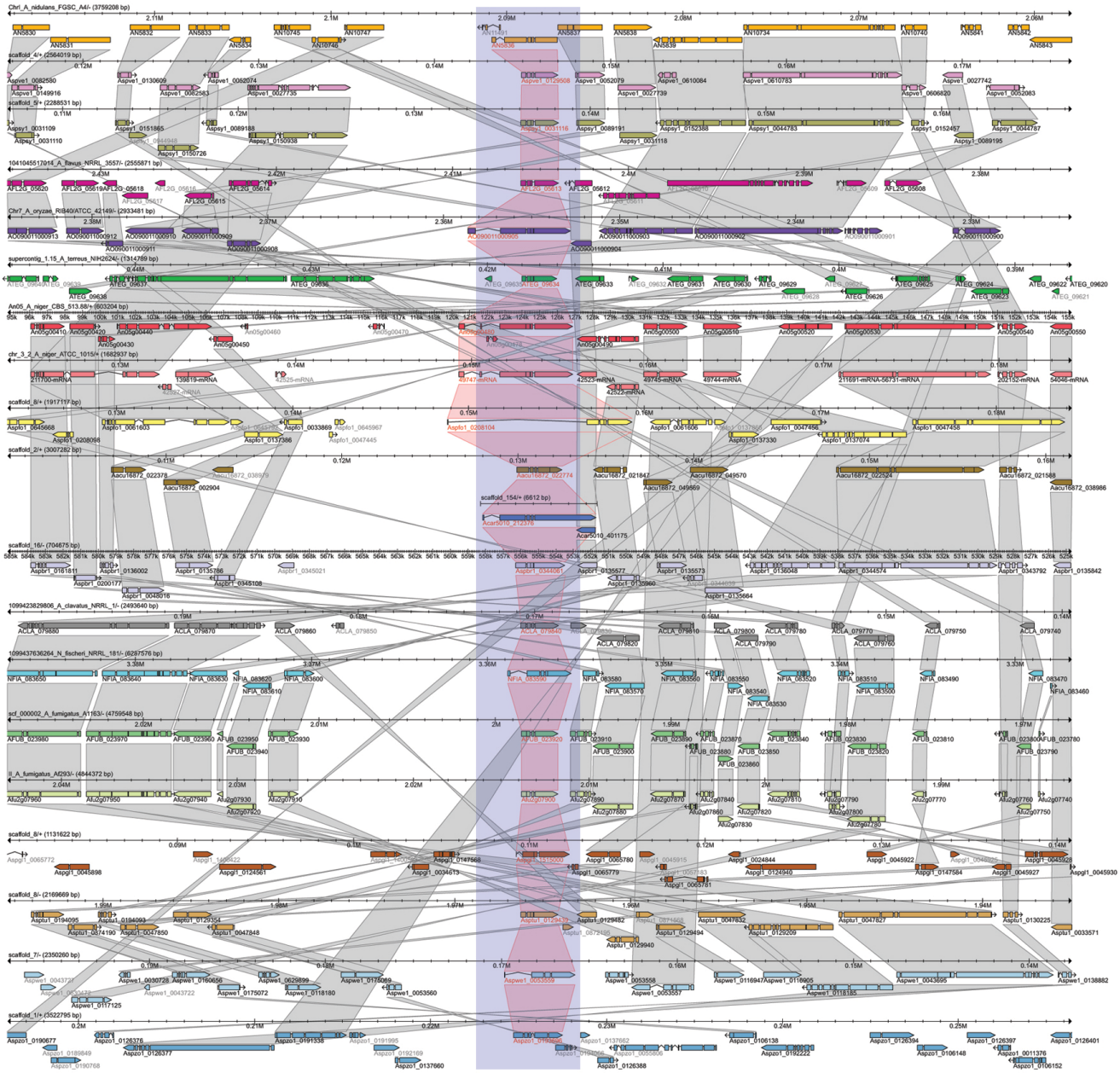


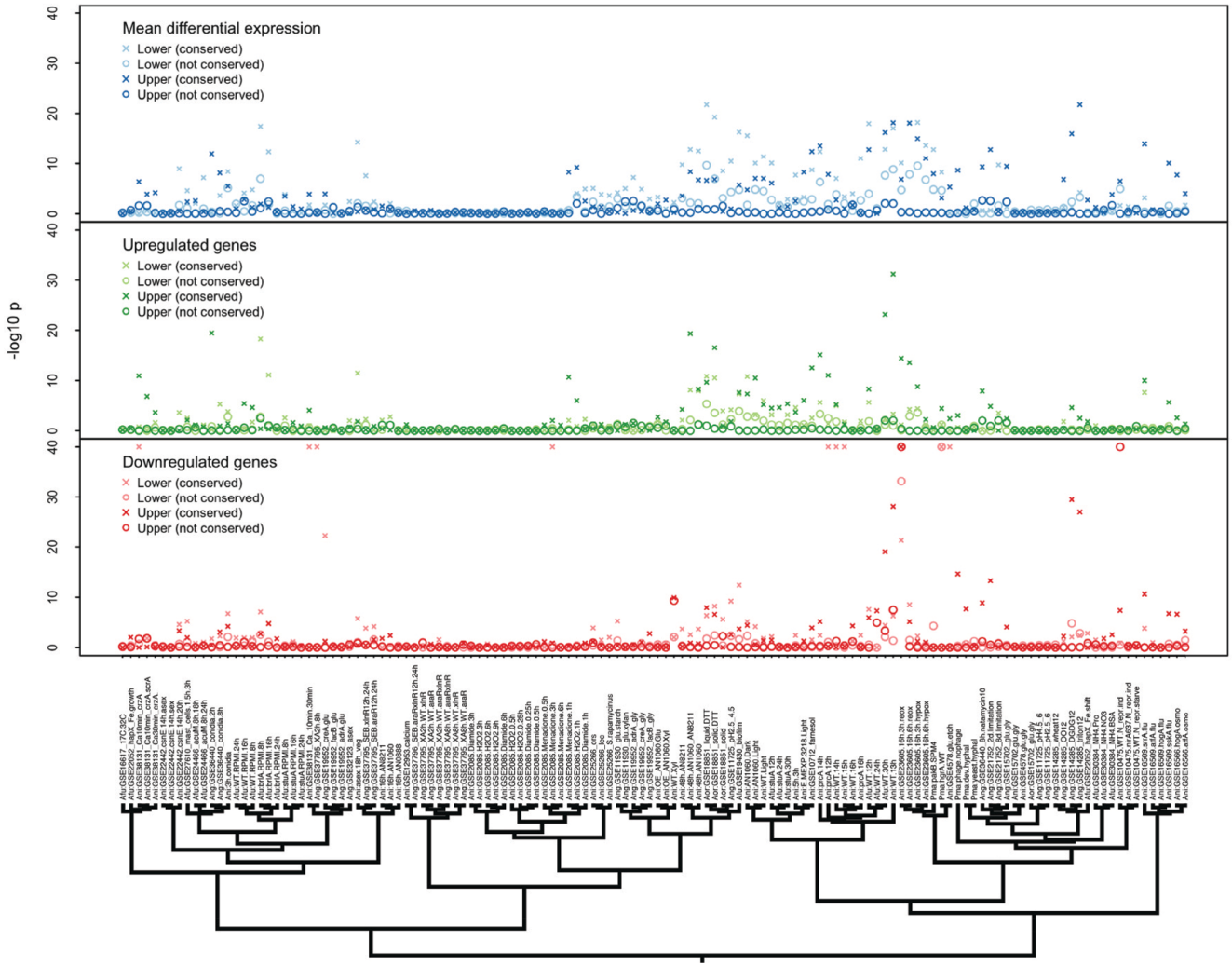
Supplementary Figure 1. Fungal phylogeny based on intergenic size for core orthologues

Pearson correlation coefficient hierarchical clustering for 1062 orthologues shared between all 17 studied species. Results recapitulate the known phylogeny approximately, with the exception of some *Aspergillus* species and the *P. marneffei*/*T. stipitatus* clade (compare with Figure 1).



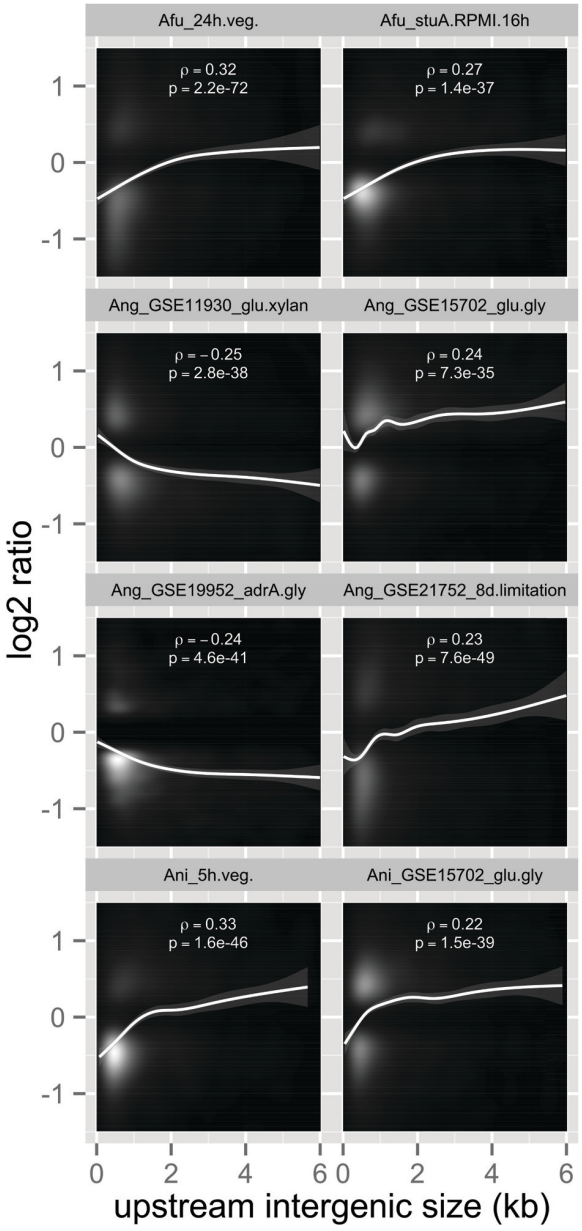
Supplementary Figure 2. Conservation of exceptional UIS size for *stuA* orthologues in *Aspergillus* species.

Synteny at the *stuA* locus (shaded in blue) for 20 strains (18 species) from the genus *Aspergillus* (AspGD Sybil browser). Plot included small ORFs classified as dubious and removed before analysis (see Methods).



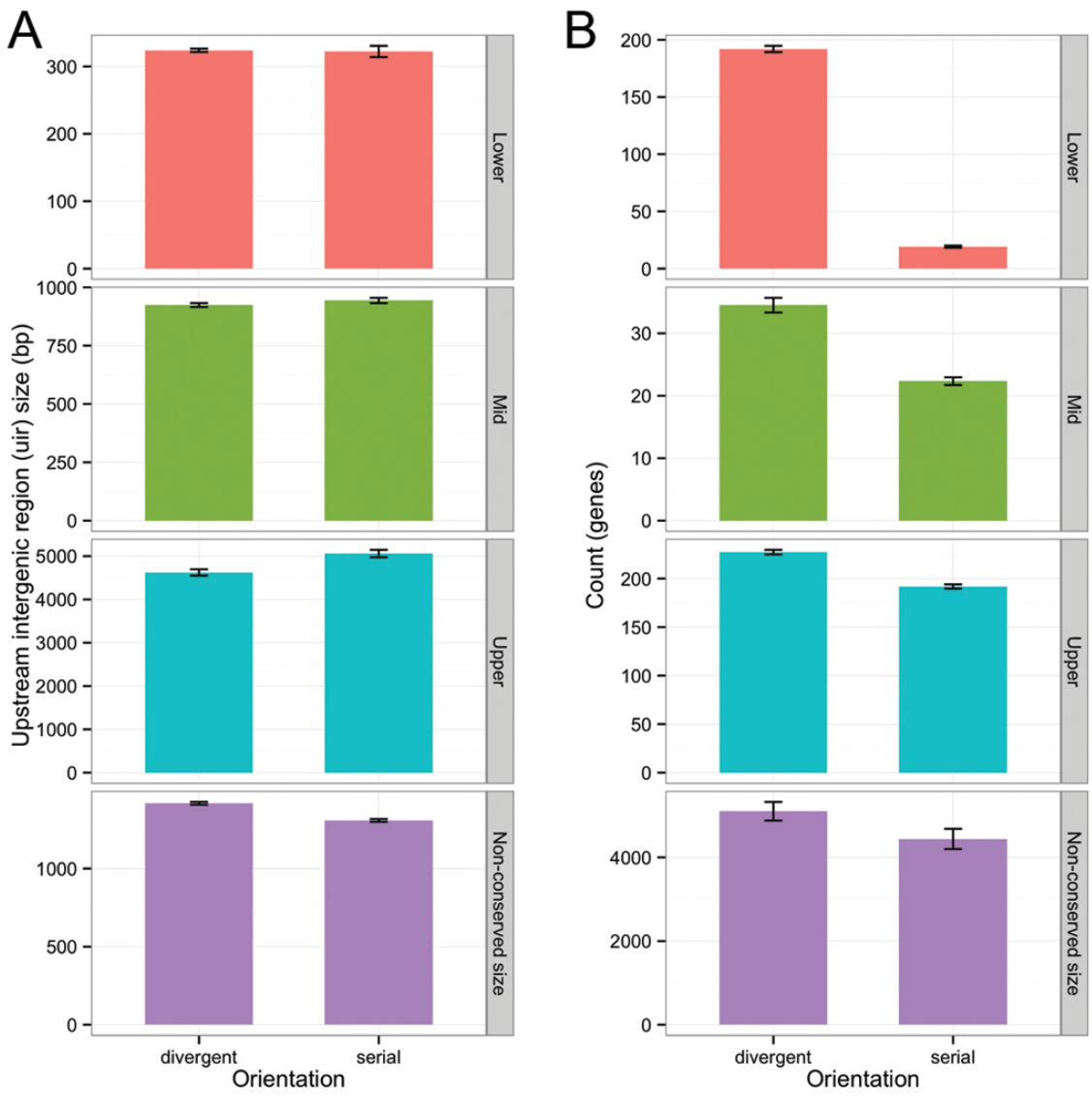
Supplementary Figure 3. Large and small UIR genes are transcriptionally promiscuous.

Plots of significance (–log10 p-value) for each test (mean differential expression, count of upregulated genes, count of downregulated genes, see Methods) and experimental condition. Note that a p-value ceiling of 10e-40 is imposed. Sample dendrogram is based on Pearson uncentred hierarchical clustering of raw data.



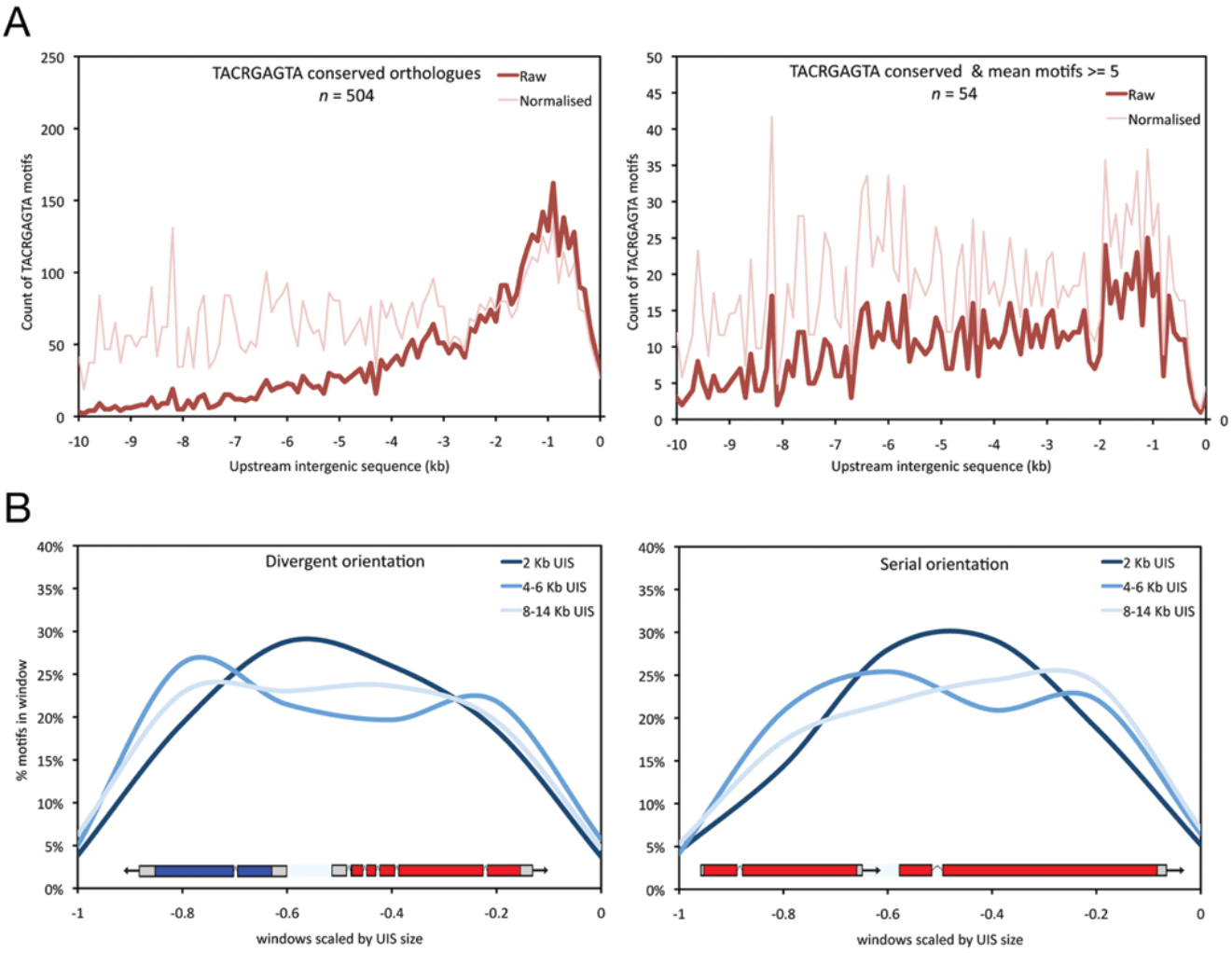
Supplementary Figure 4. Genome-wide correlation between expression and upstream intergenic size.

Density plots for individual experimental samples (ignoring orthology) showing significant correlation between expression and upstream intergenic size (Spearman's $r > 0.2$). Shading represents density of raw expression values for all genes (absolute log2 ratio > 0.2), white line is a fitted locally-weighted polynomial regression (lowess) with 95% confidence intervals. Species shown are *A. nidulans* (Ani), *A. niger* (Ang) and *A. fumigatus* (Afu). Dataset NCBI GEO accessions are shown where applicable (see Methods for further description).



Supplementary Figure 5. Relationship between upstream intergenic size and gene orientation.

A. Mean UIR size plotted for divergent and serially arranged genes in the *Aspergilli*, across first, third and fifth conserved UIR quintiles. Data for all other genes are plotted as 'Non-conserved size'. **B.** Mean count of gene orientation per UIR group across the *Aspergilli*. Error bars are standard error.



Supplementary Figure 6. A putative *cis*-regulatory motif enriched in large UIR genes.

The MEME position weight matrix for a motif enriched in conserved, large UIR genes (consensus TACRGAGTA) was used to assess genome-wide occurrence across *A. nidulans*, *A. fumigatus* and *A. oryzae* using MAST. **A.** Motifs conserved across orthologous genes were plotted by location (Kb upstream of the translation start site). Shown are raw data, and normalised data adjusted for the number of genes contributing to each data point (scale not shown). Gene-proximal bias is significant for all motifs (left, t-test for motifs within 0-2 Kb versus 2-10 Kb, p = 0.0008), but not for orthologues with multiple motifs (right, mean across orthologues >= 5, p = 0.4). **B.** Motif location was plotted separately for divergently and serially arranged genes to assess proximity bias. Location upstream of genes was scaled by intergenic size, and data were plotted for

2 Kb, 4-6 Kb and 8-14 Kb (rounded to nearest Kb) UIR genes. If promoter proximity is important for the TACRGAGTA motif then this should be particularly apparent for serially arranged genes. A weak bias for gene proximity is detected only for very large (8-16 Kb) UIR genes (corresponding to between 2 and 6 Kb). Divergently arranged genes show a broader motif distribution at all levels of UIR size, suggesting co-regulation may occur in some cases.