Divergence in a Eukaryotic Transcription Factor's co-TF Dependence Involves Multiple Intrinsically Disordered Regions Affecting Activation and Autoinhibition

Lindsey F. Snyder[1], Emily M. O'Brien[2], Jia Zhao[2#], Jinye Liang[2], Baylee J. Bruce[1], Yuning Zhang[3^], Wei Zhu[4], Thomas J. Cassier[2], Nicholas J. Schnicker[5, 6], Xu Zhou[7], Raluca Gordân[3, 4, 8, 9], Bin Z. He[1,2,*]

[1] Interdisciplinary Graduate Program in Genetics, University of Iowa, Iowa City, IA; [2] Department of Biology, University of Iowa, Iowa City, IA; [3] Department of Biostatistics & Bioinformatics, Duke University, Durham, NC; [4] Department of Molecular Genetics & Microbiology, Duke University, Durham, NC; [5] Protein and Crystallography Facility, University of Iowa, Iowa City, IA; [6] Department of Molecular Physiology and Biophysics, University of Iowa, Iowa City, IA; [7] Department of Pediatrics, Division of Gastroenterology, Hepatology and Nutrition, Boston Children's Hospital and Harvard Medical School, Boston, MA; [8] Department of Computer Science, Duke University, Durham, NC [9] Department of Cell Biology, Duke University, Durham, NC

[#] Current addresses: Laboratory of Immunophysiology, The Ragon Institute of Mass General, MIT, and Harvard; Department of Biology, Massachusetts Institute of Technology. Cambridge, MA, USA.

[^] Current addresses: Department of Genetics, Washington University School of Medicine in St. Louis, St. Louis, MO, USA.

[*] correspondence should be addressed to bin-he@uiowa.edu

Running title: Divergence in a eukaryotic TF's dependence on co-TF

ABSTRACT

Combinatorial control by multiple transcription factors (TFs) is a hallmark of eukaryotic gene regulation. Despite its prevalence and crucial roles in enhancing specificity and integrating information, the mechanisms behind why eukaryotic TFs depend on one another, and whether such interdependence evolves, are not well understood. We exploit natural variation in co-TF dependence in the yeast phosphate starvation (PHO) response to address this question. In the model yeast *Saccharomyces cerevisiae*, the main TF, Pho4, relies on the co-TF Pho2 to regulate ~28 genes. In a related yeast pathogen, *Candida glabrata*, its Pho4 exhibits significantly reduced Pho2 dependence and has an expanded target set of ~70 genes. Biochemical analyses showed *C. glabrata* Pho4 (CgPho4) binds to the same consensus motif with 3-4-fold higher affinity than ScPho4 does. A machine-learning-based prediction and yeast one-hybrid assay identified two Intrinsically Disordered Regions (IDRs) in CgPho4 that boost the activity of the main activation domain but showed little to no activity on their own. We also found evidence for autoinhibition behind the co-TF dependence in ScPho4. An IDR in ScPho4 next to its DNA binding domain was found to act as a double-edged sword: it both allows for enhanced activity with Pho2, and inhibits Pho4's activity without Pho2. This study provides a detailed molecular picture of how co-TF dependence is mediated and how its evolution, mainly driven by IDR divergence, can lead to significant rewiring of the regulatory network.

42 **INTRODUCTION**

43 Transcription factors (TFs) are the cornerstone of gene regulatory networks. In eukaryotes, TFs
44 often work collaboratively to regulate gene expression. This combinatorial control is crucial for
45 enhancing specificity, because most eukaryotic TFs recognize short and degenerate motifs –
46 typically less than 10 bps – that appear hundreds to tens of thousands of times in the genome
47 (Stewart et al., 2012). TFs with the same family of DNA binding domains (DBDs) also recognize
48 highly similar motifs (Berger et al., 2008). Despite these, TFs often bind only a fraction of their
49 motifs *in vivo* and regulate an even smaller subset of the genes they bind to (Zhou and O'Shea,
50 2011); paralogous TFs regulate distinct sets of genes (Merhej et al., 2016; Sánchez-Higueras et
51 al., 2019). The key to achieving this specific regulation is the requirement of two or more TFs to
52 jointly regulate the target genes (Avsec et al., 2021; Slattery et al., 2011; Todeschini et al.,
53 2014). Another important function of combinatorial control and TF interdependence is to allow
54 cells to integrate multiple upstream signals. In fly development, for example, a combination of
55 tissue-specific "selector" and morphogen gradient – both are TFs – precisely determine the
56 expression pattern and define the cell fate (Guss et al., 2001).

57 The importance of combinatorial control and the molecular mechanisms behind TF
58 interdependence are traditionally studied using mutations that disrupt TF interactions. For
59 instance, a missense mutation that disrupts the interaction between two cardiac TFs, GATA4
60 and TBX5, was shown to cause dysregulation of cardiac genes and lead to malformation of
61 heart tissues (Ang et al., 2016). Similarly, studies of combinatorial control evolution often focus
62 on losses (and gains) of TF interactions, like the MADS-box TF, Mcm1, and its co-TFs, MATa2
63 and MATα2 in the yeast mating type pathway (Baker et al., 2012; Britton et al., 2020). A far less
64 explored type of change in the combinatorial control is a change in the TF interdependence
65 itself, that is, when TFs involved in cooperative regulation evolved to be more or less dependent
66 on the co-TF(s). Such changes are expected to dramatically rewire the network by altering the
67 specificity and signals required for gene regulation, which could have profound implications in

68 disease and evolution. In relation to this possibility, not all eukaryotic TFs require other TFs to
69 function. The yeast TF, Gal4, regulates its target genes on its own (Traven et al., 2006). This
70 raises several intriguing and unanswered questions: if TF interdependence and combinatorial
71 control evolve, what genetic changes underlie the divergence, what aspects of the TF activities
72 are impacted, and what are the consequences on the regulatory output?

73 Intriguingly, natural variation in co-TF dependence exists in the yeast phosphate
74 starvation (PHO) response network. In the model yeast *Saccharomyces cerevisiae*, the main TF
75 of the PHO response, Pho4 (hereinafter as ScPho4), strongly depends on the co-TF, Pho2, to
76 induce 27/28 of its target genes (Zhou and O'Shea, 2011). In a related human yeast pathogen,
77 *Candida glabrata*, its Pho4 (hereinafter as CgPho4) is far less dependent on Pho2 and induces
78 twice as many genes (Fig. 1A) (He et al., 2017; Kerwin and Wykoff, 2009). The level of Pho2-
79 dependence varies quantitatively among Pho4 orthologs in other yeasts and is correlated with
80 the number of genes induced in a common genome background (He et al., 2017). This latter
81 observation is consistent with the role of combinatorial control in enhancing specificity. What
82 remains unknown is what genetic differences between Pho4 orthologs underlie the divergence
83 in co-TF dependence, and what TF activities are impacted by those variations.

84 Here, we propose two non-mutually exclusive models for the difference in Pho2-
85 dependence between ScPho4 and CgPho4 (Fig. 1B). The first "enhanced activity" model is
86 based on studies showing that ScPho4 requires Pho2 to **(1)** bind cooperatively to the target
87 gene promoters (Barbaric et al., 1998), **(2)** recruit the histone acetyltransferase (HAT) complex
88 (Nourani et al., 2004) and **(3)** help recruit general TFs and the PolII complex (Magbanua et al.,
89 1997). Under this model, we hypothesize that CgPho4 binds more tightly to DNA than ScPho4,
90 and is more capable of recruiting general transcription factors (TFs) and the PolII complex,
91 thereby becoming less Pho2-dependent. The second "autoinhibition" model is based on a study
92 suggesting that ScPho4 is auto-inhibited, and that interaction with Pho2 unmasks its activation

93 domain and allows it to function (Shao et al., 1996). This model predicts that CgPho4 either
94 lacks or has far weaker effects of the auto-inhibition and hence doesn't depend on Pho2.
95 In this study, we tested these two models by comparing the DNA binding and activation
96 abilities of the two Pho4 orthologs, and systematically swapping regions between the two Pho4s
97 in a series of 50 chimeric TFs, then quantifying their activities with and without Pho2. Our results
98 support both models as contributing to the Pho2-dependence variation. To our surprise, while
99 CgPho4 DBD binds to the same consensus motif with a 3-4-fold higher affinity than ScPho4
100 DBD, swapping DBD alone failed to yield the expected results. Instead, the differences in Pho2-
101 dependence originated primarily from differences in the Intrinsically Disordered Regions (IDRs)
102 in the two TFs that modulate both their Activation Domain (AD) and DNA binding domain (DBD)
103 activities. Therefore, our results reveal that evolution in a eukaryotic TF protein, particularly
104 through changes in the IDR, can lead to divergence in co-TF dependence, which in turn results
105 in a more than two-fold change in the size of the target network.

106 **RESULTS**

107 **Domain organization and sequence divergence between ScPho4 and CgPho4**

108 Based on genetic and biochemical studies, ScPho4 encodes the following functional domains
109 from its N- to its C-terminus (Fig. 1C): a regulatory region (R1, aa 1-42) interacting with the
110 negative regulator Pho80 (Jayaraman et al., 1994), the activation domain (AD, aa 43-99)
111 (McAndrew et al., 1998; Ogawa and Oshima, 1990), a region encoding the nuclear export and
112 import signals (NLS, aa 100-176) (Komeili and O'Shea, 1999a; O'Neill et al., 1996), a protein-
113 interaction domain interacting with both Pho80 and the co-TF, Pho2 (referred together as P2ID,
114 aa 177-242) (Hirst et al., 1994a; Jayaraman et al., 1994), and the bHLH DNA binding domain
115 (DBD, aa 243-312) (Ogawa and Oshima, 1990; Shimizu et al., 1997). Regarding ScPho4-
116 ScPho2 interaction, a previous study mapped the region in ScPho4 required for the interaction
117 to aa 200-218 (Hirst et al., 1994b). The same 18 aa stretch also contains the phosphorylation
118 site that directly controls the TF-TF interaction (Komeili and O'Shea, 1999b).

119 Both ScPho4 and CgPho4 were predicted to be mostly intrinsically disordered outside
120 the DBD, with some low confidence helices in R1 and AD (Fig. S1). CgPho4 is significantly
121 longer than ScPho4 (533 vs 312 amino acids, Fig. 1C). The NLS and P2ID regions contain four
122 functionally important phosphorylation sites targeted by the cyclin-dependent kinase complex
123 Pho80/85 (Kaffman et al., 1994; O'Neill et al., 1996). Those located in the NLS control Pho4's
124 nuclear translocation, while phosphorylation of the one in P2ID disables ScPho4 from
125 interacting with Pho2 (Komeili and O'Shea, 1999a). All five Pho80/85 targeted phosphorylation
126 sites are clearly identifiable in the alignment (Fig. S2); in fact, they were previously found to be
127 conserved in orthologs outside the Saccharomycotina subphylum (Gomes-Vieira et al., 2018).
128 Thus, while the protein length and sequence diverged significantly between the two Pho4
129 orthologs, the domain architecture and post-translational modification motifs appear to be highly
130 conserved in evolution.

131 **CgPho4 binds the same consensus motif with a higher affinity compared to ScPho4**

132 To test the enhanced activity model, we first compared the DNA binding ability of the two Pho4
133 orthologs. First, we asked if the two Pho4 orthologs recognize the same DNA sequence. The N-
134 terminal stretch of the first α-helix in the bHLH domain, known as the basic region, contains the
135 residues determining sequence specificity. All four residues known to recognize the nucleotide
136 bases and five additional residues that recognize the phosphate backbone in ScPho4 (Shimizu
137 et al., 1997) are all conserved in CgPho4 except for an R252K change (number based on
138 ScPho4, Fig. 2A). Next, we used our published Chromatin-Immunoprecipitation (ChIP-seq)
139 data for both Pho4 orthologs to identify their respective motifs (He et al., 2017) (Materials and
140 Methods). Consistent with the amino acid sequence conservation, the results showed that they
141 bind the same E-box motif "CACGTG", with no obvious differences in the flanking nucleotides
142 (Fig. 2B). These motifs were based on a relatively small number of ChIP peaks (74 for ScPho4
143 and 118 for CgPho4). To comprehensively characterize and compare the binding preferences of
144 the two proteins, we applied Protein Binding Microarray (PBM) to map the binding landscape in

145 the entire 7-mer space ("universal" PBM, Fig. 2C). The result revealed a lack of divergence in
146 their sequence preference (Pearson's r = 0.89). To complement the short oligo length in the
147 universal PBM and further examine differences in the flanking base preference, we designed a
148 second "genome context" library, which includes 36-bp sequences centered on ChIP-identified
149 binding sites for both Pho4's in their native genome along with the flanking nucleotides (Gordân
150 et al., 2013). Similar to the uPBM, the genome context PBM revealed no evidence for binding
151 specificity differences between ScPho4 and CgPho4 (Fig. S3, Spearman's ρ = 0.84 and 0.86 for
152 sequences containing the consensus E-box or non-consensus variants, respectively, Materials
153 and Methods). We conclude that ScPho4 and CgPho4 have the same sequence specificity.
154 Next, we tested the hypothesis that CgPho4 binds more tightly than ScPho4 does. To do
155 so, we purified the DBD region of both Pho4s (Materials and Methods) and measured their
156 binding affinity to a 17-bp oligo based on the *S. cerevisiae PHO5* promoter with the consensus
157 E-box motif "CACGTG". Biolayer Interferometry (BLI) measurements showed that CgPho4 DBD
158 binds the consensus motif >3 times more tightly than ScPho4 DBD does (K_D = 5.2 nM vs 1.2 nM
159 for ScPho4 and CgPho4, respectively; Student's t-test for log K_D difference $P < 0.01$, Fig. 2D,
160 E). We confirmed this K_D difference using Electrophoretic Mobility Shift Assay (K_D = 5.5 nM vs
161 1.9 nM for ScPho4 and CgPho4, respectively; Fig. S4). We conclude that CgPho4 recognizes
162 the same E-box motif and binds with a higher affinity than ScPho4. It is worth noting that this
163 observed difference in affinity, while statistically significant, is modest ($\Delta\Delta G$ = -0.87 kcal/mol,
164 based on the mean BLI measurements, assuming a temperature of 25°C). Its potential impact
165 on gene regulation depends critically on the effective concentration of Pho4 in the nucleus.
166 **CgPho4 encodes two additional Activation Enhancer Domains (AEDs)**
167 We next tested the hypothesis that CgPho4 has increased transactivation potential compared to
168 ScPho4 under the enhanced activity model. We first predicted regions with activation potential
169 in both Pho4 orthologs using PADDLE, a Convolutional Neural Network trained on 150
170 activation domains (ADs) from 164 TFs in *S. cerevisiae* (Sanborn et al., 2021). PADDLE

171 recovered the experimentally identified AD in ScPho4 between aa 60-102, which contained the
172 9aaTAD motif previously described as the minimum residues required for activation (Piskacek et
173 al., 2007) (Fig. 3A, orange triangle). No other regions with significant activation potential were
174 predicted in ScPho4. By contrast, three regions in CgPho4 were predicted to have activation
175 capabilities, including one corresponding to the AD in ScPho4, which was predicted to have
176 strong activity (Z-score > 6, Fig. 3B). Two additional regions, one overlapping R1 and the other
177 spanning the NLS and P2ID, were predicted to have medium activation strength (Z-score > 4).
178 We refer to these two regions as E1 and E2, respectively from here on. Interestingly, the best
179 match to the 9aaTAD motif pattern in CgPho4 was found in E2 rather than in the AD (Fig. 3B).
180 To determine the activity of the predicted regions with activation potential in both Pho4s,
181 we set up a yeast one-hybrid (Y1H) assay, in which each candidate region was fused to the
182 Gal4 DBD and its activation potential was measured by a genome-integrated *GAL1*pr-mCherry
183 reporter in *S. cerevisiae* (Fig. 3C, Materials and Methods). To see if fusing the candidate region
184 with Gal4 DBD created a new activation domain, we applied PADDLE to all constructs and
185 observed no peaks either in the Gal4 DBD or in the connecting regions (Fig. S5). The yeast
186 one-hybrid result confirmed that both ScPho4 and CgPho4 ADs were able to activate the
187 reporter above the background level (7- and 5.6-fold, Holm-Bonferroni corrected $P < 0.01$). We
188 were able to further localize the required residues for activation in ScPho4's AD to a region of
189 32 aa centered on the 9aaTAD motif (Fig. 3A, C). Neither E1 nor E2 from either Pho4 activated
190 the reporter on its own (Fig. 3C). This was surprising for CgPho4's E1 and E2, since both had a
191 medium strength Z-score prediction, and the latter also contained a match to the 9aaTAD (Fig.
192 3B). We hypothesized that E1 and E2 in CgPho4 could enhance the activity of the main AD. To
193 test this, we first fused ScE1 and CgE1 to their respective AD, and found that only the latter
194 resulted in a significant enhancement in the activation potential (Fig. 3D). Next, we fused CgE1
195 or CgE2 in both orientations to the minimal AD of ScPho4 (ScAD_9aa). We found that indeed,
196 both regions were able to significantly enhance the activity of the ScAD_9aa in an orientation-

197 independent manner (Fig. 3D). It is worth noting that this enhancement activity does not involve
198 the endogenous ScPho4 or ScPho2. ScPho4 is inactive in the phosphate replete condition in
199 which the Y1H experiment was conducted. To test the involvement of ScPho2, we repeated the
200 assay in a host lacking *pho2*. No significant difference in the boosting effects of either CgE1 or
201 CgE2 was observed, although CgE1's effect is slightly lower without Pho2 (Fig. S6).
202 In summary, we found that CgPho4 encodes two Activation Enhancer Domains (AEDs),
203 both of which are in the Intrinsically Disordered Region (IDR). The AEDs have little activation
204 capability on their own but can significantly enhance the activation potential of the AD of either
205 of the Pho4 orthologs. Since these two AEDs are present in CgPho4 but not in ScPho4, we
206 hypothesize that they contribute to CgPho4 being less dependent on Pho2.
207 **A dual fluorescence reporter assay accurately measures the activity of Pho4 chimeras**
208 **with and without Pho2 for dissecting the divergence in co-TF dependence**
209 So far, we compared DNA binding and activation potential between CgPho4 and ScPho4 by
210 isolating individual regions and testing them either *in vitro* or in a synthetic *in vivo* system. While
211 they support the enhanced activity model, it remains unclear whether and how they contribute to
212 divergence in Pho2-dependence in the native Pho4 protein context. They also do not test or rule
213 out the autoinhibition model, which necessitates intramolecular interactions. To answer these
214 questions, we divided ScPho4 and CgPho4 into five corresponding parts (Fig. 1C, 4A), with
215 breakpoints chosen to be on the edge of well-aligned regions to avoid breaking known or
216 predicted functional domains and secondary structures. Boundaries for all regions used in this
217 and other experiments can be found in Table 1. We created all 2^5=32 combinations as well as
218 28 additional ones with alternative breakpoints (Fig. 4A). All constructs were C-terminally tagged
219 with mNeon to quantify the protein levels, and were expressed from the native *ScPHO4*
220 promoter and UTRs (Fig. 4B). We also created two *S. cerevisiae* host strains, in which the
221 *ScPHO5* CDS was replaced by an mCherry reporter, and *pho80* was knocked out so that all
222 Pho4 chimeras were constitutively nuclear localized (Kaffman et al., 1994; O'Neill et al., 1996).

223 One of the two host strains had *pho2* knocked out. For each chimera, we measured its *PHO5pr-*
224 mCherry and Pho4-mNeon levels using flow cytometry in both hosts. While mCherry and
225 mNeon levels for the same Pho4 construct varied between experiments, the ratios between the
226 two were consistent and characteristic of the specific construct (Fig. 4C, D). We therefore
227 defined the activity of a Pho4 chimera as the ratio between the median fluorescence intensity
228 (MFI) of mCherry and mNeon, which we will refer to as A_{PHO2} or $A_{pho2\Delta}$ from hereon.

229 **The two Activation Enhancer Domains (AEDs) in CgPho4 increased the activity of the**
230 **chimeric TFs but were insufficient to make them Pho2-independent**

231 The chimeric Pho4 constructs exhibited varied A_{PHO2} and $A_{pho2\Delta}$ values (Fig. 5A). Interestingly,
232 some showed higher A_{PHO2} than either ScPho4 or CgPho4. To identify the potential genetic basis
233 for the divergence in Pho2-dependence, we first calculated the activity difference when one or
234 two regions of ScPho4 were replaced with their counterpart(s) from CgPho4. The results were
235 plotted as two heat maps separated based on the presence of Pho2 (Fig. 5B). For example, a
236 chimera with its NLS from CgPho4 (NLS:Cg) and the rest from ScPho4 led to estimates of the
237 activity difference between NLS:Cg and NLS:Sc on ScPho4 background with or without Pho2.

238 Based on our previous results, we expected R1, NLS, and DBD in CgPho4 to enhance
239 the activity of the chimera with and without Pho2. In particular, the first two encode CgE1 and a
240 majority of CgE2_9aa, which we showed enhanced the activity of the main AD (Fig. 3); DBD:Cg
241 is expected to increase the binding ability of the chimera (Fig. 2). We indeed found R1:Cg and
242 NLS:Cg to enhance A_{PHO2} on ScPho4's background (Fig. 5C, "R1, AD, NLS" group); contrary to
243 our expectation, however, DBD:Cg reduced the chimera's activity with Pho2 (Fig. 5C, "bHLH").
244 Also unexpectedly, R1:Cg and NLS:Cg had a much smaller effect on $A_{pho2\Delta}$ (Fig. 5B bottom, 5C
245 right), suggesting that despite their ability to increase A_{PHO2}, the chimeras were still dependent
246 on Pho2. In fact, none of the 1- or 2-region swaps led to large increases in $A_{pho2\Delta}$. It is worth
247 noting that this result also argues against a classic lock-and-key model for autoinhibition, which
248 we expect to be broken by swapping one of the regions involved.

249 To quantitatively examine the contribution of R1, AD, and NLS regions from CgPho4 and
250 whether they interact non-additively (epistasis), we fit a linear model to the data to estimate the
251 main and interaction terms for the regions, i.e., $Y = X_0 + R1 + AD + NLS + R1{:}AD + R1{:}NLS +$
252 $AD{:}NLS + R1{:}AD{:}NLS$. In this model, the first term represents the activity of ScPho4, the next
253 three terms the main effect of each CgPho4 region (":Cg" omitted for brevity), and the rest are
254 interaction terms. We found that R1:Cg and NLS:Cg each had significant, positive effects on
255 their own on both A_{PHO2} and $A_{pho2\Delta}$; although the magnitude was much smaller for $A_{pho2\Delta}$ (Fig.
256 5D, Holm-Bonferroni corrected $P < 0.05$). Both also had a significant and positive interaction
257 term with AD:Cg on A_{PHO2} (corrected $P < 0.05$ for R1:AD and AD:NLS); the estimates trended in
258 the same direction but were not significant for $A_{pho2\Delta}$ (Fig. 5D). One explanation for the observed
259 epistasis may be that the two CgPho4 AEDs work more efficiently with AD:Cg than with AD:Sc.
260 However, it could also be explained by the disruption of the native conformation and function of
261 the "interacting" regions resulting in a lower activity in the species-mixed constructs. Note that to
262 minimize such effects, we designed the breakpoints to avoid any predicted secondary structures
263 (Fig. S1, S2).
264 Since the NLS:Cg region encodes both the NLS/NES and a key part of the second AED
265 (CgE2_9aa), we asked which of these two mechanisms was responsible for its positive effect on
266 A_{PHO2}. To test the possibility that the enhanced A_{PHO2} was due to a stronger nuclear localization
267 activity of NLS:Cg, we performed fluorescence microscopy to quantify the concentration of the
268 nuclear-localized Pho4 proteins and the ratio of nuclear vs total Pho4 proteins in six constructs
269 that bear either NLS:Sc or NLS:Cg. No significant difference between the two groups was found
270 (Fig. S7, F-test $P > 0.1$). We thus conclude that the positive effect of NLS from CgPho4 on A_{PHO2}
271 was mainly due to its ability to boost activation.
272 We also examined the effect of swapping CgPho4's DBD. On its own, swapping DBD:Cg
273 decreased both A_{PHO2} and $A_{pho2\Delta}$ compared to ScPho4 (Fig. 5B), suggesting that DBD:Cg may
274 be incompatible with one or multiple regions in ScPho4.

275 In summary, we found that R1:Cg and NLS:Cg increased the activity of the chimera with
276 Pho2 but were insufficient to remove the Pho2 dependence alone or in combination. In the
277 presence of Pho2, R1:Cg and NLS:Cg both showed positive epistasis with AD:Cg on ScPho4's
278 background. Lastly, no 1- or 2-region swaps from CgPho4 into ScPho4 increased $A_{pho2\Delta}$ to the
279 level of CgPho4. Together, we conclude that CgPho4's two AEDs can increase the activity of
280 the chimera in the presence of Pho2, although they fail to restore the activity without Pho2 to
281 CgPho4's level, suggesting that additional mechanisms are at play. CgPho4's DBD, alone or in
282 combination with R1 and the NLS region, did not contribute to increased activity.

283 **A double-edged sword: Pho2 Interaction Domain (P2ID) in ScPho4 allows for enhanced**
284 **activity with Pho2 but restricts its activity without Pho2**

285 In the 1- and 2- region swap heatmap (Fig. 5B), the P2ID showed a puzzling pattern: swapping
286 CgPho4's P2ID (P2ID:Cg) into the ScPho4 background had a dominant negative effect on A_{PHO2}
287 while offering little increase in $A_{pho2\Delta}$. This suggests that P2ID:Sc is essential for ScPho4's
288 function. Conversely, swapping ScPho4's P2ID (P2ID:Sc) into CgPho4 increased A_{PHO2} beyond
289 that of CgPho4 (18 to 22) but reduced $A_{pho2\Delta}$ (17 to 4.8). This suggests that P2ID:Sc is a key
290 factor to Pho2-dependence. To investigate the unique property of the P2ID further, we plotted
291 all chimeras based on their A_{PHO2} and $A_{pho2\Delta}$ values (Fig. 6A). Strikingly, the chimeras fell into
292 three distinct groups based on the identity of their P2IDs. The first group had P2ID from
293 CgPho4, and their activities were not dependent on Pho2, falling along the diagonal line. The
294 second group had P2ID from ScPho4. These had strong activity with Pho2 and low to no activity
295 without Pho2, like ScPho4 does. The third group included additional chimeras with mixed P2ID
296 regions, and they filled the intermediate space between the first two groups. Notably, many
297 chimeras in the second group showed higher A_{PHO2} than ScPho4. Most of them contained the
298 R1 and NLS regions from CgPho4, consistent with our results above showing that these two
299 regions enhanced the activity of the Pho4 chimeras in the presence of Pho2 (Fig. 5A, B).

300 The evidence above suggests that P2ID:Sc has two effects: allowing for collaborative
301 regulation with Pho2 and restricting Pho4's activity without Pho2. Can these two functions be
302 separated? To answer this question, we compared the mixed P2ID chimeras to those with
303 whole P2IDs from either Pho4. We found that swapping P2ID:Sc into CgPho4 resulted in a 1.2-
304 fold increase in A_{PHO2} and a 71% reduction in $A_{pho2\Delta}$ (Fig. 6B, row 2 vs 3). Interestingly,
305 swapping just the second half of P2ID:Sc (aa 205-242) into CgPho4 resulted in a ~40%
306 reduction in A_{PHO2} and 75% reduction in $A_{pho2\Delta}$. By contrast, just swapping the first half of
307 P2ID:Sc (aa 177-204) resulted in 1.76-fold increase in A_{PHO2} and maintained the same level of
308 $A_{pho2\Delta}$ (1.1-fold). From this, we deduced that the first half of P2ID:Sc mainly functions to interact
309 with Pho2 while the second half appears to limit Pho4's activity without Pho2. The same trend
310 was observed in another set of chimeras with lower activities (Fig. 6B, rows 6-9).
311 Several chimeras had very low activities even with Pho2 (A_{PHo2} < 3.6, or 20% of A_{PHO2} for
312 ScPho4). Most of these nonfunctional chimeras have P2ID:Cg and DBD:Sc (Fig. 6C, rows 3-7).
313 We hypothesize that DBD:Sc requires P2ID:Sc and Pho2 to function in the context of the full
314 length Pho4. This is contrary to the conventional view that DBDs can function on their own,
315 which is supported by our own *in vitro* results (Fig. 2). Nonetheless, we reasoned that if the
316 above hypothesis is correct, putting P2ID:Sc back by inserting it in between P2ID:Cg and
317 DBD:Sc should rescue the non-functional chimeras in the presence of Pho2. That is what we
318 observed (Fig. 6C, rows 8-12), with some of the chimeras even exceeding the A_{PHO2} of ScPho4
319 and CgPho4. However, all of them still required Pho2. These results support the above
320 hypothesis, showing that the dual-functional P2ID is essential for ScPho4 to function.
321 Given that chimeras with P2ID:Cg have equal activities with and without Pho2 (Fig. 6A), we
322 asked if CgPho4 still physically interacts with Pho2 in *S. cerevisiae*. Using the yeast two-hybrid
323 assay, we were unable to detect an interaction between CgPho4ΔDBD (aa 2-463) and a region
324 of ScPho2 known to interact with multiple TFs, including ScPho4 (Bhoite et al., 2002) (Fig. 6D).
325 By contrast, ScPho4ΔDBD (aa 2-250) was able to interact with the same region of ScPho2 as

326 previously found (Fig. 6D). This is consistent with our observation above, where chimeras with
327 P2ID:Cg and lacking the region from ScPho4 mediating Pho2-interaction cannot be enhanced
328 by Pho2 (Fig. 6A).

329 In summary, our chimera dissection revealed three key regions behind the difference in
330 co-TF dependence. Notably, all three are IDRs. Among them, a region adjacent to the DBD in
331 ScPho4 (P2ID:Sc) functions as a double-edged sword: it both allows Pho4 to gain activity with
332 Pho2's help and restricts it when Pho2 is absent. We showed that these two functions are
333 encoded by physically separate parts, offering a path for determining their respective
334 mechanisms of actions. CgPho4 lacks the ability to interact with and use Pho2's help. Instead,
335 the two AEDs we identified earlier – both in IDR – conferred higher activity independent of
336 Pho2. This, combined with the lack of the autoinhibition by P2ID:Sc, makes CgPho4 as active
337 as ScPho4 and not dependent on the co-TF.

338 **DISCUSSION**

339 Combinatorial control plays crucial roles in eukaryotic gene regulation. Mutations disrupting TF
340 interactions can cause dysregulation and disease (Ang et al., 2016). However, how mutations
341 can alter TF interdependence itself, whether in disease or evolution, is less understood. In this
342 study, we investigated the molecular basis for natural variation in co-TF dependence in the
343 yeast phosphate starvation (PHO) response. We found three key differences between two
344 orthologous Pho4s with varying dependence on the co-TF Pho2 (Fig. 7): **(1)** DNA Binding
345 Affinity: CgPho4 binds the same consensus DNA motif with 3-4-fold higher affinity than ScPho4;
346 **(2)** Activation potential**:** CgPho4 has two unique activation enhancing domains (AEDs) that
347 increase the activation potential of both Pho4s. **(3)** Autoinhibition: ScPho4 contains an IDR next
348 to its DBD that both allows it to interact with Pho2 to gain enhanced activity, and inhibits its
349 activity in the absence of Pho2. Therefore, our results support both the enhanced activity model
350 and the autoinhibition model for the difference in Pho2-dependence between Pho4 orthologs.

351 Among the three differences, the contribution of CgPho4's two AEDs to the TF's activity
352 and dependence on Pho2 is well supported by the yeast one-hybrid data and chimeric Pho4
353 results (Fig. 3, 5). We are not aware of existing examples or proposed mechanisms for such a
354 phenomenon. One hypothesis for how AEDs work is that they are weaker ADs with the same
355 biochemical activities, i.e., recruiting cofactors through protein-protein interactions. As such,
356 they are not sufficient for activation by themselves but can increase the activity of the nearby AD
357 (the effect can be non-additive). Alternatively, AEDs may affect the conformation of the AD and
358 have no activity on their own. Further tests by synthetic constructs and biochemical assays,
359 such as co-IP followed by mass-spec, will provide mechanistic insight into how AEDs function.
360 By contrast, the significance of the binding affinity differences between the two Pho4s
361 remains unclear (Fig. 2E). Binding kinetics predicts that if the nuclear concentrations of both
362 Pho4 are much higher than their K_D, a 3-4-fold difference in affinity would have little impact.
363 Conversely, if the nuclear concentration of Pho4 is close to its K_D, the same difference could
364 significantly affect gene induction. Existence of the second scenario is supported by a study
365 showing that nuclear ScPho4 levels are much lower at intermediate phosphate concentrations
366 than in extreme starvation conditions, leading to differential binding and induction of ScPho4's
367 targets (Ming Yip et al., 2023). In our chimeric Pho4 experiments, replacing ScPho4's DBD with
368 that from CgPho4 reduced Pho4's activity rather than increasing it both with and without Pho2
369 (Fig. 5). It seems to suggest that the binding affinity difference has no functional impact *in vivo*.
370 However, it is worth noting that we assayed the activity of chimeric Pho4s in the *pho80*Δ
371 background where all Pho4 proteins are constitutively inside the nucleus at near-maximal levels,
372 hence not allowing us to test the effect of the affinity difference at a lower level of nuclear Pho4.
373 Future studies will need to measure Pho4 activities at varying nuclear concentrations to test the
374 above hypothesis.
375 Surprisingly, we found that ScPho4's DBD requires both ScPho4's P2ID and Pho2 for
376 full activity (Fig. 6C), challenging the view that TFs are modular, where DBD are expected to

377 function independently. Although ours and others' studies confirmed that ScPho4 DBD can bind
378 DNA *in vitro* (Fig. 2) and even *in vivo* when overexpressed on its own (Donovan et al., 2023),
379 they may not fully reflect its physiological activity, which depends on the nuclear concentration,
380 nucleosomal context, and interactions with other TF regions and cofactors. Further experiments
381 are needed to resolve the paradox and re-examine the assumption of TF modularity.

382 Only one of the three functional differences identified between the two Pho4 orthologs is
383 in a structured region. The two AEDs and the dual-functional P2ID:Sc are both within IDRs.
384 IDRs are abundant in eukaryotic TFs, with over 80% having at least one such region (Liu et al.,
385 2006). Approximately 75% of ScPho4 and CgPho4 were predicted to be IDR (Fig. S1). While
386 most studies of TF function and evolution focus on the structured regions like the DBD, our work
387 joins a small number of studies highlighting the significance of IDR divergence in TF evolution
388 (Hsu et al., 2021; Lynch et al., 2011). On one hand, IDRs play crucial roles in nearly all aspects
389 of a TF's function, including recruiting cofactors, interacting with co-TFs, contributing to binding
390 specificity and forming molecular condensates (Brodsky et al., 2020; Ferrie et al., 2023, 2022; Ji
391 et al., 2024; Krois et al., 2018; Mindel et al., 2023). On the other hand, IDRs evolve much faster
392 than structured regions, potentially offering more raw variation for natural selection to act on.
393 Our current understanding of how IDR evolves and affects TF function is limited by challenges
394 in aligning their sequence and studying their functions. Progress in experimental techniques like
395 phase separation (Ji et al., 2024) and large language models trained on protein sequences offer
396 new opportunities to investigate TF IDR's function and evolution (Rives et al., 2021).

397 Our study focused on dissecting the genetic basis for Pho2-dependence variation in two
398 Pho4 orthologs. We previously showed that this trait varies across Pho4 orthologs, with reduced
399 Pho2-dependence potentially evolving independently in more than one lineage (He et al., 2017).
400 We wonder whether the IDR-associated divergence identified above also correlated with the
401 level of Pho2-dependence more broadly. Preliminary analyses of the activation potential and

402 P2ID length across eight Pho4 orthologs with different Pho2-dependence levels supported this
403 speculation (Fig. S8).

404 Lastly, our study illustrates a novel mechanism for the evolution of combinatorial control
405 in eukaryotes. Unlike previous research that focused on the gain and loss of TF interactions in
406 gene regulatory networks (Tuch et al., 2008; Wapinski et al., 2010), we show that TF's
407 dependence on co-TF itself can evolve. This divergence can lead to significant rewiring of the
408 regulatory network, as seen with Pho4, where reduced co-TF dependence correlates with an
409 expanded target gene network (He et al., 2017). This contrasts with the dominant pattern seen
410 in the literature on combinatorial control evolution: the Johnson lab, for example, showed that
411 changes in TF combinations in the yeast mating type pathway altered the mode of regulation
412 but maintained the overall network output (Baker et al., 2012; Britton et al., 2020; Nocedal and
413 Johnson, 2015; Tsong et al., 2006). While cases of TF interdependence evolution are still rare,
414 it is interesting to compare our example with Gain-of-Function mutations in kinases: in the proto-
415 oncogene ABL, a fusion with another gene called BCL makes the merged protein independent
416 of the upstream and downstream activators, which drives excessive cell proliferation and leads
417 to Chronic Myeloid Leukemia (Sherbenou and Druker, 2007). Whether similar changes in TF
418 dependence on co-TFs lead to misregulation and in turn causes diseases or novel phenotypes
419 is an extremely interesting question for future studies.

420 In summary, our study provides a detailed molecular picture of how co-TF dependence
421 is mediated and how it evolves, particularly through IDR changes. Further exploration of both
422 questions is essential for understanding gene regulation and regulatory evolution in eukaryotes.

423 **MATERIALS AND METHODS**

424 Breakpoints used for the chimeric Pho4 constructs and individual regions are listed in Table 1.
425 Plasmids and strains are listed in Tables 2 and 3. Computational and statistical analysis scripts

426 performed in this study are available at https://github.com/binhe-lab/E013-Pho4-evolution, which
427 will be archived using Zenodo and minted with a DOI at the time of publication.

428 **Bioinformatic analyses of Pho4 orthologs**

429 Pho4 ortholog sequences were from the Yeast Gene Order Browser (http://ygob.ucd.ie/), and
430 were aligned using ProbCons (Do et al., 2005) via JalView's Web Service (Troshin et al., 2018;
431 Waterhouse et al., 2009). The alignment was manually edited to align the five Pho80/85 motifs
432 (O'Neill et al., 1996). Secondary structures for ScPho4 and CgPho4 were predicted using the
433 PSIPRED 4.0 (Buchan and Jones, 2019). 9aaTAD motifs were predicted using the webapp
434 https://www.med.muni.cz/9aaTAD/ (Piskacek et al., 2007). For ScPho4, one match was found
435 using the moderately stringent pattern, located at aa 75-83; for CgPho4, one match was
436 identified using the most stringent pattern, at aa 270-278, with three more using the moderately
437 stringent patterns (aa 20-28, 24-32, 282-290). Among the latter three, aa 24-32 had a lower %
438 match at 67% vs 83% for the others.

439 **Identifying Pho4 binding motifs from ChIP-seq data**

440 Chromatin Immunoprecipitation (ChIP-seq) for ScPho4 and CgPho4 were previously performed
441 in *S. cerevisiae*, with both Pho4 expressed from the same endogenous ScPho4 locus with
442 native regulatory sequences (He et al., 2017). ChIP identified peaks for both Pho4s were
443 downloaded from the supplementary files of the above publication. The sequences under each
444 peak were extracted from *S. cerevisiae* genome (sacCer3, NCBI refseq assembly
445 GCF_000146045.2), which were submitted to the "peak-motifs" tool without control sequences
446 on the RSAT Fungi server (https://rsat.france-bioinformatique.fr/fungi/) (Thomas-Chollier et al.,
447 2011). The top motif was reported for each Pho4 ortholog.

448 **Protein expression and purification for DNA binding assays**

449 All recombinant proteins were expressed using pET-11a (Sigma #69436-3) based vectors in
450 BL21(DE3) *E. coli* cells. Vector maps are available upon request. The DNA binding domain

451 (DBD) constructs included ScPho4 DBD-6xHIS (aa 236-312) and CgPho4 DBD-6xHIS (aa 452-
452 533) cloned downstream of the T7 promoter in pET-11a. The transformed bacterial strains were
453 grown overnight in LB + 100 ug/mL ampicillin. The overnight culture was diluted to OD 0.1 and
454 the cells were grown to OD 0.6 and induced with 1mM IPTG for 2 hours. Induced cells were
455 collected by centrifugation at 6000 rpm at 4ºC for 35 minutes, and were snap frozen in liquid
456 nitrogen, then stored at -80ºC until purification. Our preliminary experiments showed that the
457 protein of interest was largely in the insoluble fraction. Therefore, we performed a refold protocol
458 on the Ni-NTA column. Briefly, frozen pellets were resuspended in 1xPBS, pH7.4 with cOmplete
459 EDTA-free protease inhibitor (Sigma #11836170001) and sonicated for 45 cycles of 1 sec on / 2
460 sec off repeated 3 times at 50% power to lyse the cells. Sonicated samples were centrifuged at
461 35k rpm for 30 minutes. The supernatant was discarded, and the pellet was resuspended in the
462 solubilization buffer (20 mM TrisBase, 0.5 M NaCl, 5 mM imidazole, 5.5 M Guanidine
463 Hydrochloride, 1 mM 2-mercaptoethanol pH 8.0) and stirred at ~25ºC for an hour. The
464 solubilized pellet was centrifuged at 35k rpm for 30 minutes and the supernatant was filtered
465 and loaded onto a 5mL Ni-NTA column. The column was washed with a urea buffer (20 mM
466 TrisBase, 0.5 M NaCl, 20 mM imidazole, 5.5 M Urea, 1 mM 2-mercaptoethanol pH 8.0). Then,
467 the protein was refolded with a reverse urea gradient from buffer A (20 mM TrisBase, 0.5 M
468 NaCl, 20 mM imidazole, 5.5 M Urea, 1 mM 2-mercaptoethanol pH 8.0) to buffer B (20 mM
469 TrisBase, 0.5 M NaCl, 20 mM imidazole, 1 mM 2-mercaptoethanol pH 8.0) on a BioRad FPLC.
470 The refolded protein was eluted off of the Ni-NTA column using a gradient from buffer C (20 mM
471 TrisBase, 0.5 M NaCl, 20 mM imidazole, 1 mM 2-mercaptoethanol pH 8.0) to buffer D (20 mM
472 TrisBase, 0.5 M NaCl, 500 mM imidazole, 1 mM 2-mercaptoethanol pH 8.0). Fractions
473 containing the protein of interest were identified by gel electrophoresis, pooled, and diluted with
474 a no salt buffer (25 mM Na_2HPO_4 pH 7.0, 0.5 mM THP) and run on a Heparin column
475 equilibrated with low salt buffer (25 mM Na_2HPO_4 pH 7.0, 0.15 M NaCl, 0.5 mM THP). Protein
476 was eluted with a gradient from low to high salt buffer (25 mM Na_2HPO_4 pH 7.0, 1.5 M NaCl, 0.5

477 mM THP). Protein-containing fractions were concentrated using a 3 kDa cutoff Amicon
478 centrifugal filter (Sigma UFC8003) and loaded onto a Superdex 75 size exclusion column
479 equilibrated with a storage buffer (25 mM Na_2HPO_4 pH 7.0, 0.5 M NaCl, 0.5 mM THP).
480 Fractions containing the expected size products were collected, analyzed by gel
481 electrophoresis, and stored in the storage buffer at 4ºC.
482 N-GST-CgPho4 full length was constructed for the Protein Binding Microarray assay.
483 The corresponding N-GST-ScPho4 purification has been described before (Zhang et al., 2021).
484 Both proteins were grown and lysed as described in (Zhang et al., 2021). Briefly, BL21 *E. coli*
485 cells containing the constructs were induced at OD600 0.8 with 1mM IPTG and collected by
486 centrifugation after 3 hours of induction at 30ºC. Pellets were snap frozen in liquid nitrogen and
487 stored at -80ºC until purification. Cells were lysed with rLysozyme (Millipore 71110) for 20
488 minutes at room temp in the lysis buffer (1xPBS, pH7.4) with the cOmplete protease inhibitor
489 tablet and 1 mM PMSF. The protein was run on a GST column equilibrated with the lysis buffer
490 and eluted with a gradient to buffer B (50mM Tris, 10mM glutathione, pH 8). GST-CgPho4
491 fractions containing the protein were pooled and loaded onto a heparin column equilibrated with
492 low salt buffer (25mM Na_2HPO_4 pH 7.4, 150mM NaCl, 0.5mM THP) and eluted with a gradient
493 to high salt buffer (25mM sodium phosphate dibasic pH 7.4, 1.5M NaCl, 0.5mM THP). Fractions
494 containing the protein were pooled and concentrated with a 30 kDa cutoff Amicon centrifugal
495 filter (Sigma UFC9030) and loaded onto a Superdex 200 size exclusion column equilibrated with
496 the storage buffer (25mM HEPES pH 7.4, 500mM NaCl, 0.5mM THP). Protein was run on an
497 SDS-PAGE and pure fractions were pooled and concentrated. 10% glycerol was added before
498 snap freezing and storage at -80ºC.
499 **Universal Protein Binding Microarray (uPBM)**
500 The uPBM was performed following the PBM protocol as described in (Berger and Bulyk, 2009).
501 GST-ScPho4 and GST-CgPho4 (full length) were prepared as described above. Both proteins
502 were assayed at a final concentration of 1 uM as determined by the optical absorbance. An

503 8x15k array was used to assay all possible 9-mers, from which a robust 7-mer enrichment score
504 is derived. The non-parametric enrichment score, or E-score, is invariant to differences in the
505 concentration of the proteins used in the assay, and thus are suitable for comparisons of relative
506 affinities between arrays. E-score ranges from -0.5 (lowest enrichment) to +0.5 (highest
507 enrichment). Scores greater than 0.35 correspond to specific TF-DNA binding (Berger and
508 Bulyk, 2006; Gordân et al., 2013). Data analysis was performed using custom Perl scripts as
509 described in (Berger and Bulyk, 2009).

510 **Genomic Context Protein Binding Microarray (gcPBM)**

511 All probes in the DNA library are 60 bp in length with 24 bp complementary to the primer used
512 for double stranding and 36 bp of genomic region centered on the E-box motif or its variant. The
513 library contains 1) 5711 *S. cerevisiae* genomic regions with putative Pho4 binding sites; 2) 150
514 negative controls from the *S. cerevisiae* genome not specifically bound by Pho4 (Zhang et al.,
515 2021); 3) 4000 DNA sequences used in previous MITOMI experiments to calibrate the binding
516 affinities (Maerkl and Quake, 2007); 4) 150 genomic regions from *C. glabrata* that contain Pho4
517 consensus and nonconsensus binding sites (He et al., 2017); 5) 100 probes from the library of
518 sequences we tested with Biolayer Interferometry; and 6) 150 negative controls from the *C.*
519 *glabrata* genome not bound by CgPho4. We used the NNNNGTG, CACNNN, and GTGNNN
520 libraries from Maerkl and Quake (2007) in our gcPBM design. The MITOMI and BLI probes
521 required the addition of random flanks to maintain the 36bp length;10 different flanking
522 sequences were generated for each sequence. Each probe is represented by six replicates,
523 including three replicates in each orientation and were distributed randomly across the array.
524 The custom 8x60k (8 chambers, 60,000 DNA spots per chamber) was synthesized by Agilent
525 Technologies. The gcPBM was performed following the PBM protocol as described in (Gordân
526 et al., 2013; Zhang et al., 2021). The log transformed median intensity of the 6 replicate probes
527 were used for comparisons. Because the log signal intensities are not directly comparable
528 between the two Pho4 proteins, which were assayed on separate arrays, we used Spearman's

529 rank correlation coefficient to quantify the level of concordance in their sequence preference.
530 Any difference in preference should result in a change in the ranks within each Pho4's data.

531 **Biolayer interferometry (BLI)**

532 A library of 17-bp dsDNA was used for the assay. The consensus probe had the sequence
533 "CTAGTCC**CACGTG**TGAG", with the E-box motif bolded, and was identical to the DNA used in
534 the crystal structure of ScPho4's DBD (Shimizu et al., 1997). Nine half-site variants, including
535 "AACGTG, TACGTG, … CAAGTG, ACGGTG, CATGTG" were constructed on this background.
536 For each probe, the complementary ssDNA oligos were synthesized by Integrated DNA
537 Technologies (IDT). One of the two probes was biotinylated on the 5' end. To anneal them into
538 dsDNA, 1 pmol (5 uL of 200 uM) of the biotinylated oligo was mixed with 2 pmol (10 uL of 200
539 uM) of the complementary, unmodified strand in the Nuclease-Free Duplex Buffer (30 mM
540 HEPES, pH 7.5; 100 mM potassium acetate). The mixture was heated to 95ºC and then left at
541 room temperature for it to cool down. Annealed probes were stored at -20ºC until use.

542 ScPho4 DBD and CgPho4 DBD were purified and stored as described above. Protein
543 quality was checked weekly for signs of degradation using Dynamic Light Scattering (DLS) on a
544 DynaPro NanoStar instrument. Before each experiment, the concentration of the protein prep
545 was measured in triplicates using a NanoDrop instrument. The mean concentration was used to
546 prepare the protein dilutions and calculate K_D.

547 BLI experiments were performed on an Octet RED 96 instrument at 30ºC with 1000 rpm
548 shaking. For each protein-probe pair, eight streptavidin (SA) biosensors were hydrated for 15-20
549 min at room temperature in the 1x kinetics buffer (1x PBS pH 7.4, 0.01% BSA, 0.002% Tween-
550 20, 0.005% Sodium azide). A black 96 well flat-bottom plate was loaded with experimental
551 components also diluted in the 1x kinetics buffer. To begin the experiment, biosensors were
552 equilibrated in a 1x kinetics buffer for 60 seconds to reach the baseline. Seven biosensors were
553 then submerged in 35 nM biotinylated annealed DNA for 30 seconds, while one biosensor was
554 submerged in the 1x kinetics buffer as a no-DNA control. Biosensors were then submerged in 1

555 ug/ml biocytin for 60 seconds to block any empty streptavidin pocket on their surface, before
556 being dipped back into a 1x kinetics buffer for 60 seconds for a baseline measurement. Loaded
557 and blocked biosensors were then submerged into a gradient of protein concentrations
558 calculated for each probe based on the K_D. To obtain reliable measurements, we use a
559 concentration range that spans 10x to 1/10 of the K_D values. Biosensors stayed in the protein
560 solution for 900-1000 seconds or until equilibrium was reached.
561 Data analysis was performed in the ForteBio Data analysis v11 software. After
562 subtracting the background and aligning the y-axis, the processed data were subjected to either
563 a steady state analysis or a kinetic curve fitting. For steady state analysis, the equilibrium-level
564 signal from each biosensor was plotted against the protein concentration, from which K_D was
565 calculated. Kinetic curve fitting was done using a one-site specific binding with Hill slope model
566 as implemented in the Data analysis v11 software. The latter more effectively fit the BLI data
567 and thus were utilized for most of the analysis. For the consensus sequence, we found the
568 kinetic curve-based estimates for both ScPho4 and CgPho4 to have higher variance, likely due
569 to the fast kinetics not adequately captured by the model (Weeramange et al., 2020). Steady-
570 state analysis gave consistent mean K_D as the kinetic analysis did, but resulted in lower
571 variance, and was used instead. Two to four replicates were performed for each 17bp DNA
572 library sequence, using at least two independent protein preps.
573 **Electrophoretic Mobility Shift Assay (EMSA)**
574 IR700 labeled 17bp consensus DNA (same as the consensus DNA for BLI above, Integrated
575 DNA technologies) was diluted to 0.1 nM and mixed 1:1 with a 2x dilution series of the DBD of
576 interest starting at 32nM in binding buffer (20mM Tris-HCL, pH 8, 150mM KCl, 10% glycerol,
577 5mM MgCl2, 1mM EDTA, and 1mM DTT). The mixture was incubated for 1 hour at 4ºC. A 10%
578 native PAGE gel in 1x Tris Glycine buffer (2.5 mM Tris pH 8.3, 19.2 mM glycine) was prerun at
579 200V for 20 min. The DNA/protein mixture was then loaded onto the gel and ran for 25-35
580 minutes. The gel was imaged using the Odyssey FC imager in the 700 nm channel. To estimate

581 K_D, the unbound band in each protein-containing lane was quantified using the ImageStudio
582 software (with background subtraction) and divided by the no protein control. This value was
583 then subtracted from 1 and plotted against the protein concentrations. A nonlinear curve fitting
584 was performed in Prism v10.2.1 using the one site specific binding model. K_D estimates and the
585 standard deviations were reported.

586 **Yeast media and growth**

587 Yeast cells were grown in Yeast extract-Peptone Dextrose (YPD) medium or Synthetic
588 Complete (SC) medium, using Yeast Nitrogen Base without amino acids (Sigma Y0626)
589 supplemented with 2% glucose and amino acid mix. Phosphate starvation medium was made
590 using Yeast Nitrogen Base with ammonium sulfate, without phosphates, without sodium chloride
591 (MP Biomedicals, 114027812) and supplemented to a final concentration of 2% glucose, 1.5
592 mg/ml potassium chloride, 0.1 mg/ml sodium chloride and amino acids, as described previously
593 (He et al., 2017). Phosphate concentration in the medium was measured using a Malachite
594 Green Phosphate Assay kit (Sigma, MAK307).

595 **Yeast strain and plasmid construction**

596 The hosts containing the endogenous *PHO5*pr-mCherry reporter were constructed by replacing
597 the endogenous coding sequence with mCherry using CRISPR/Cas9. Briefly, guide RNAs
598 targeting *PHO5* were designed in Benchling (https://www.benchling.com) and cloned into
599 bRA89 (AddGene 100950), a plasmid that contains the Cas9 protein and gRNA scaffold (Anand
600 et al., 2017). Homology arms to the *PHO5* 5'UTR and 3' UTR were added onto an mCherry
601 donor DNA using PCR. The donor DNA and plasmid were co-transformed into the host using
602 the standard LiAc transformation protocol. Transformants were selected for the CRISPR
603 plasmid using hygromycin resistance and screened with PCR. Positive clones were validated
604 using Sanger sequencing and fluorescence microscopy. Pho2 was then knocked out using a
605 *HIS3* cassette with homology arms to the *PHO2* 5'UTR and 3'UTR.

606 The chimeric Pho4 plasmid library was constructed using Golden Gate. The fragments
607 of CgPho4 and ScPho4 were PCR amplified using Phusion Flash polymerase (Thermo
608 Scientific F548S) with unique overhangs for Golden Gate. They were then assembled and
609 inserted into a pRS315-based backbone that contains the Sc*PHO4* promoter, a C-terminal in-
610 frame mNeon tag followed by the Sc*PHO4* 3' UTR and terminator. Primers were designed using
611 the NEBridge Golden Gate assembly tool. The inserts were verified with PCR and Sanger
612 sequencing and then transformed into the yeast hosts using either the standard LiAc protocol
613 (Gietz and Schiestl, 2007) or the Zymo yeast transformation kit (Zymo research T2001) and
614 plated onto SD-leu media.

615 **Flow cytometry**

616 Cells were inoculated into a 96 deep-well plate (Fisher Scientific 07-200-700) and grown
617 overnight in SD-Leu or SC medium supplemented to final concentrations of 0.13 mg/ml adenine
618 and 0.1 mg/ml tryptophan to reduce autofluorescence (Lam et al., 2008). Cultures were diluted
619 in the morning to an OD600 of 0.15 and grown to an OD600 of 0.6. These cells were directly
620 subjected to flow cytometry at a flow rate of 25 uL/min on an Attune NxT flow cytometer
621 (ThermoFisher) fitted with an autosampler. Data were collected using the Attune NxT software
622 v3.1 for FSC, SSC and the appropriate fluorescence channels. Calibration beads (Spherotech
623 RCP305A) were run routinely to ensure that experiments from different days were comparable.
624 Pho4-mNeon was measured in the BL1 channel with 488 nm excitation and 510±10 nm
625 emission; mCherry was measured in the YL2 channel using 561 nm excitation and 615±25 nm
626 emission. Voltages for each channel were set by the brightest sample and negative control so
627 the sample signals were between 10^2-10^5. Events were gated based on FSC-H / SSC-H to
628 remove non-cells, then on the FSC-W / FSC-H to isolate singlets (Fig. S9). At least 10,000
629 gated events were collected per sample. Each strain was measured at least three times, and on
630 two different dates. A nonfluorescent strain was included in every experiment for subtracting the

631 autofluorescence. Further gating and analyses were performed in R using the FlowClust and
632 FlowCore packages. Detailed analysis scripts are available in the project github repository.

633 **Fluorescent microscopy for Pho4 nuclear localization**

634 Six Pho4 constructs were chosen, with three bearing NLS:Sc (SSSSS, CCSCC, SSSCC) and
635 three bearing NLS:Cg (SSCSS, CCCCC, SSCCS). They were transformed into either a *PHO2*
636 wild type or *pho2*Δ background. Yeast cells were grown to mid-log phase in SD-Leu and fixed
637 using 4% paraformaldehyde for 10 minutes, washed twice with 1xPBS before DAPI staining.
638 4',6-diamidino-2-phenylindole (DAPI, Sigma, D9542) was diluted in the respective medium and
639 added to the fixed cells at a final concentration of 10 ug/mL. Cells were incubated with DAPI in
640 the dark for 30 minutes before DAPI was removed and cells were washed twice with 1xPBS.
641 Fixed cells were mounted on a 1.2% agarose pad on a depression slide for image. Fluorescent
642 imaging was performed on a Leica epifluorescence microscope. A 405 nm laser was used for
643 DAPI excitation and 430-550 nm for emission. For Pho4-mNeON, 488nm was used for
644 excitation and 530 nm for emission. A bright field image was also taken to show the cell
645 boundaries. A total of three images were recorded for each field of view.

646 Microscopy analysis was performed in ImageJ (v 1.53). To quantify the nuclear fraction
647 of Pho4 proteins, ten cells were randomly selected for each construct in each host background,
648 and the experiment was repeated two times, resulting in a total of 15 or 20 cells quantified in
649 *PHO2* and *pho2*Δ backgrounds, respectively. Cell boundaries were manually traced using the
650 bright field channel, and nucleus using the DAPI channel. Both were added to the ROI manager,
651 and the raw integrated density (sum of pixel values) was quantified for both areas. The ratio of
652 nuclear vs whole cell integrated density was used for plotting and statistical analysis.

653 **Yeast one-hybrid and yeast two-hybrid**

654 Indicated regions from CgPho4, ScPho4, and ScPho2 were PCR amplified and cloned into
655 pGBD-C3 (James et al., 1996) containing the Gal4 DBD using Gibson Assembly. Plasmids were
656 verified with PCR and Sanger sequencing. The *GAL1*pr-mCherry reporter was created using a

657 pRS306 based integrative plasmid digested with StuI and inserted into the *ura3* locus of a *gal4Δ*
658 *gal80Δ S. cerevisiae* strain. The reporter was tested using flow cytometry according to the
659 protocol above using only the red (YL2) channel. Plasmids were then transformed into this yeast
660 strain using the standard LiAc method (Gietz and Schiestl, 2007) or the Zymo yeast
661 transformation kit (Zymo research T2001) and selected for on SD-trp. For the yeast two-hybrid,
662 the Gal4 activation domain (AD) and Gal4 DNA binding domain (DBD) fusion plasmids were
663 constructed by PCR amplifying the indicated regions and cloning into pGBD-C3 and pGAD-C3
664 (James et al., 1996) using Gibson Assembly. The plasmids were co-transformed and selected
665 for on SD -leu -trp. Positive colonies were patched onto fresh plates and grown at 30ºC for 24
666 hours before replica plating onto SD -leu -trp -ade -his to test for interactions. Plates were
667 imaged after 45-50 hours.

668 **Statistical analyses**

669 All replicates are biological. For yeast strains, the same strain was grown and measured in
670 separate vials. Independent transformants were tested for the flow cytometry host strains and
671 randomly selected constructs and were found to generate consistent results. For recombinant
672 proteins, the same constructs were independently transformed into the bacteria and separate
673 batches of purified proteins were used for the measurements.

674 <u>*Binding affinity comparison by BLI*</u> (Fig. 2): K_D estimates from either steady state or kinetic curve
675 fitting analyses were log10 transformed and a Student's t-test was used to compare ScPho4 vs
676 CgPho4 DBDs against the same 17-bp DNA. Raw two-sided *P*-values and Holm-Bonferroni
677 corrected *P* values for the 10 tests were reported; <u>*Activation*</u> (Fig. 3): median fluorescence
678 intensity (MFI) for the *GAL1pr*-mCherry reporter was recorded from flow cytometry for each
679 sample. For Fig. 3C, each Gal4 DBD fusion construct was compared to the background level.
680 Because the host (with *GAL1pr*-mCherry but no Gal4 DBD plasmid) and Gal4 DBD alone
681 showed a similar level of low MFI, we combined them as the reference group. A linear model
682 was fit in R with the following command `lm(MFI ~ Genotype)`, where "Genotype" is a factor

683 representing the constructs. This model estimates a common standard deviation for all
684 constructs and tests the significance of each construct against the reference group with a two-
685 sided t-test. The raw *P*-values were corrected for multiple testing using the Holm-Bonferroni
686 procedure. Constructs with a corrected *P*-value <= 0.05 were considered significant. For Fig.
687 3D, the first two groups were tested using a two-sample t-test, while the third involved multiple
688 levels, and were tested as described above for Fig. 3C. A Holm-Bonferroni correction was
689 applied to the raw *P*-values for all six tests together. *Epistasis between regions R1, AD, and*
690 *NLS of CgPho4* (Fig. 5D): to determine the main and interaction terms in these swaps from
691 CgPho4 on ScPho4's background, we fit a linear model, $Y = X_0 + R1 + AD + NLS + R1{:}AD +$
692 $R1{:}NLS + AD{:}NLS + R1{:}AD{:}NLS$ in R, using the command `lm(A ~ R1 * AD * NLS)`, where "A"
693 is either A_{PHO2} or $A_{pho2\Delta}$, and the independent variables are coded as ScPho4 = 0, CgPho4 = 1.
694 The raw *P*-values were adjusted for multiple testing using the Holm-Bonferroni procedure. *Split*
695 *P2ID* (Fig. 6B): The same procedure as applied to the activation region test above (Fig. 3) was
696 applied to the split P2ID swaps. Two sets of four chimeras were chosen, each with a reference
697 construct having P2ID:Cg, one with the entire P2ID swapped for the ScPho4 version, and two
698 with the first or second half of P2ID swapped for the ScPho4 version. The latter three were
699 compared to the reference using a linear model `lm(MFI ~ Genotype)`, for A_{PHO2} and $A_{pho2\Delta}$
700 separately. Holm-Bonferroni correction was applied to the two sets combined (6 tests in total),
701 again separately for A_{PHO2} and $A_{pho2\Delta}$. A corrected $P < 0.05$ was considered significant. *Nuclear*
702 *fraction of Pho4 chimeras* (Fig. S7): The ratio of the sum of pixel values in the GFP channel
703 inside the nucleus vs the whole cell was treated as the response variable, and the identity of the
704 NLS region ("Cg" vs "Sc") as the predictor. First, a two-way ANOVA was performed with the
705 formula `lm(Nuc_frac ~ NLS + Host)`, where Host is either *PHO2* or *pho2*Δ. Second, given that
706 ratios are often non-normally distributed, we performed the non-parametric Kruskal-Wallis test
707 with the command `kruskal.test(Nuc_frac ~ NLS)` separately in the *PHO2* and *pho2*Δ hosts. The

ANOVA F-test *P*-value (0.13) was reported in the results. The two *P*-values for the second test were 0.42 (*PHO2*) and 0.36 (*pho2*Δ).

DATA AVAILABILITY

Source data for figures in this paper are available at https://doi.org/10.5281/zenodo.14501732. Raw microscopy images for quantifying Pho4 nuclear concentration are available from the corresponding author upon request. No restrictions apply to any of the data generated in this study.

CODE AVAILABILITY

Scripts for figures and statistical tests are available at https://doi.org/10.5281/zenodo.14501732.

ACKNOWLEDGEMENT

We would like to thank Dr. Miles Pufall's lab for teaching us EMSA. We thank Dr. Jan Fassler for sharing many yeast plasmids for the yeast one- and two-hybrids. Dr. Kim Harold generously shared the TET-on inducible promoter plasmid with us. Drs. Jan Fassler, Miles Pufall, Craig Ellermeier and Todd Washington all critically read the thesis chapter by LFS, on which this manuscript is based. We thank Kyle Malcolm for helping with the early development of the BLI assay. Christian Weinrich discovered the P2ID's effect on Pho2-dependence during his rotation. We thank Dr. Yann Vanrobaeys for helping with an initial analysis using PADDLE. We would like to acknowledge the use of resources at the Protein and Crystallography Facility within the Carver College of Medicine at the University of Iowa and thank Lokesh Gakhar, Zhen Xu, and Devin Reusch for assistance with protein purification and BLI assays.

FUNDING

LFS is supported by the NIH Predoctoral Training Grant T32GM008629. BJB is supported by the NIH Predoctoral Training Grant T32GM1454441. BZH is supported by NIH R35GM137831 and a startup fund from the University of Iowa.

732 **FIGURE LEGENDS**

733 **Fig. 1 Different co-TF (Pho2) dependence between orthologous TF (Pho4) in yeast**
734 **species.** (A) Cartoon summary of previous work: *S. cerevisiae* Pho4 (ScPho4) depends on the
735 co-TF, Pho2, for DNA binding and activation of Phosphate Starvation (PHO) Response genes
736 (top row), while Pho4 from *C. glabrata* (CgPho4) can bind and activate without the co-TF. See
737 text for references. (B) Two non-mutually exclusive models to explain the divergence in Pho2-
738 dependence. (C) The tiles in the middle show the conservation scores at aligned positions of the
739 two Pho4 orthologs; the box diagrams represent ScPho4 and CgPho4, with gray lines
740 connecting matching residues to the alignment tiles. Five regions were delineated based on
741 knowledge about ScPho4: R1 = regulatory, interact with Pho80; AD = activation domain; NLS =
742 nuclear import and export signals; P2ID = Pho2-interaction domain (also includes a region
743 contacting Pho80); DBD = basic Helix-Loop-Helix DNA binding domain. The lines and cylinders
744 above and below the region diagrams show predicted α helices in each protein, with lighter
745 color corresponding to lower prediction confidence. Prediction was done using PSIPRED 4.0.
746

747 **Fig. 2 CgPho4 DBD recognizes the same motif as ScPho4 DBD and has a higher affinity**
748 **towards the consensus DNA.** (A) Alignment of the DBD region between the two Pho4
749 orthologs. Top: sequence logo for the bHLH domain (PROSITE PS50888). Middle: coordinates
750 above are based on ScPho4. Filled and open triangles indicate residues in ScPho4 DBD
751 making either base or phosphate backbone contacts (Shimizu *et al*., 1997). Bottom: predicted α-
752 helices in both Pho4 DBDs. (B) WebLogo motifs for ScPho4 and CgPho4 are derived from
753 ChIP-seq peaks. (C) Universal Protein Binding Microarray (uPBM) comparing the binding of full
754 length ScPho4 and CgPho4 against all possible 9-mers. A rank-order based E-score is plotted
755 for all 7-mers, where a score greater than 0.35 indicates specific TF-DNA binding (Gordân *et al*.,
756 2013). Red dots in the top right corner represent oligos with the “CACGTG” motif. (D) Binding to
757 the 17bp oligo containing a consensus E-box motif is measured by Biolayer interferometry (BLI).

758 Fractional binding is plotted vs the concentration of purified ScPho4 or CgPho4 DBD. One of the
759 representative binding curves is shown here. (E) Dissociation constants (K_D) are determined
760 from the binding curves and shown as dots (n = 5). The bars and the red dots are the means
761 and the red lines the 95% confidence intervals from bootstrapping. Two-sided Student's t-test
762 was performed on the log transformed K_D. A Wilcoxon rank-sum test on raw K_D similarly yielded
763 $P < 0.01$.
764
765 **Fig. 3 CgPho4 encodes two Activation Enhancer Domains (AEDs) that can enhance the**
766 **activation potential of the main AD.** (A, B) PADDLE prediction of activation potential in
767 ScPho4 and CgPho4. Predictions were made in 53 aa blocks and plotted using the middle of
768 each block as the x coordinate. As a result, the first and last 26 aa positions have no scores. Z-
769 scores > 4 and 6 indicate regions with medium and strong activation potential, respectively. The
770 five regions in each protein are labeled as in Fig. 1. Orange triangles mark the 9aaTAD motif
771 matches. Below the plots are diagrams of the Gal4DBD fusion constructs used to test the
772 activation potential, with their boundaries labeled on the right. C) Activation potential of each
773 region was tested in the context of a Gal4 DBD fusion driving a *GAL1pr*-mCherry reporter.
774 Shown are the mean (bar) and individual data points (n=3, except for "host", for which n = 6).
775 "host" has the mCherry reporter but no plasmid; "Gal4_DBD" has the reporter and the Gal4 DBD
776 alone. ANOVA was used to compare the Gal4 DBD fusions to the two negative controls
777 combined; an asterisk indicates significant difference at a 0.05 level after Holm-Bonferroni
778 correction. D) CgE1 and CgE2_9aa can enhance the activation capability of ScAD_9aa in an
779 orientation-independent manner. An asterisk indicates significance difference at a 0.05 level by
780 a t-test between the indicated groups after Holm-Bonferroni correction.
781
782 **Fig. 4 A dual fluorescence reporter system for accurately quantifying the activity of Pho4**
783 **chimeras with and without Pho2.** (A) ScPho4 and CgPho4 were divided into five

784 corresponding parts. Names of each region were based on the annotations for ScPho4.
785 Predicted alpha helices were drawn as cartoons outside of the box diagrams. All $2^5 = 32$
786 combinations (2 were the original proteins) were made in the main chimera set, and an
787 additional set of chimeras were constructed to test specific hypotheses. (B) A diagram showing
788 the design of the dual fluorescence reporter system. The chimeric Pho4 fragments were cloned
789 downstream of the endogenous *ScPHO4* promoter and were tagged at the C-terminus with an
790 in-frame mNeon fluorescent protein, followed by the endogenous *ScPHO4* 3'UTR and
791 terminator. The plasmid contains a CEN/ARS element, resulting in stable inheritance and low
792 biological variation. (C) The *PHO5pr-mCherry* levels correlate with the ScPho4-mNeon levels
793 between biological replicates. The linear regression line is shown in blue and a 95% confidence
794 band as a gray ribbon. (D) A_{PHO2} and $A_{pho2\Delta}$ values for ScPho4 and CgPho4 across 6 different
795 days of experiments showing consistent results (n = 3 for all groups except ScPho4 with Pho2,
796 where n = 12)
797
798 **Fig. 5. R1 and NLS of CgPho4 confer stronger activity than their counterparts in ScPho4.**
799 (A) Pho4 chimeras (black), endogenous ScPho4 (green) and CgPho4 (blue) were plotted based
800 on their activity with and without Pho2. The solid line indicates equal activity (independent of
801 Pho2). (B) Heatmap showing the difference in activity between the CgPho4 region(s) and the
802 counterpart of ScPho4 measured in the ScPho4 background. The upper half shows the
803 difference with Pho2 and the lower half without Pho2. The triangles (e.g., a) show the difference
804 between single regions while the squares on the top show the differences when swapping two
805 regions (see key to the right). Red color indicates a higher activity of the region from CgPho4
806 than the counterpart in ScPho4 and blue color indicates decreased activity. (C) A subset of the
807 chimeras were shown with their A_{PHO2} and $A_{pho2\Delta}$ values. Bars represent the means; lines are
808 95% confidence intervals based on bootstrap; plus signs are individual data points (n>=5),
809 which were not shown for the endogenous ScPho4 and CgPho4 (n>=36). Vertical dotted lines

810 equal the endogenous ScPho4 level. (D) Epistasis analysis of CgPho4 R1, AD and NLS. A
811 linear model was fit to the data with all two and three-way interactions. Shown are estimates of
812 the individual and interaction terms (bar and point) for the CgPho4 region(s). Line ranges
813 indicate the standard errors of the estimates. Dark gray indicates the term is significant at 0.05
814 level after Holm-Bonferroni correction.
815
816 **Fig. 6 P2ID in ScPho4 allows for enhanced activation with Pho2 via physical interaction**
817 **with the co-TF.** (A) Scatter plot of all chimeric Pho4 constructs based on their A_{PHO2} and $A_{pho2\Delta}$
818 values, colored by the identity of their P2ID region. Mixed P2ID uses an additional breakpoint
819 that separates the region into two halves (Fig. S2). (B) Comparing the effects of swapping the
820 whole P2ID from ScPho4 into CgPho4 vs individual halves. Bars represent the means; lines are
821 95% confidence intervals based on bootstrap; plus signs are individual data points (n>=5),
822 which were not shown for the endogenous ScPho4 and CgPho4 (n>=36). Two sets were tested
823 (rows 2-5 and 6-9). Two-sided Student's t-tests were performed as indicated by the bars and
824 lines to the right. Holm-Bonferroni correction was applied to the raw *P*-values for both sets,
825 separately for A_{PHO2} and $A_{pho2\Delta}$. An asterisk indicates corrected $P < 0.05$. (C) Testing the hypothesis
826 that ScPho4 DBD requires the P2ID:Sc and Pho2 to be fully functional. Plot designs are the
827 same as in (B). Compared with the chimeras with P2ID:Cg + P2ID:Sc, which are all non-
828 functional even with Pho2, adding back P2ID:Sc ("4+5") rescued the activity with Pho2, but not
829 without. (D) Pho4-Pho2 physical interaction was tested with a yeast two-hybrid assay. The bait
830 and target regions were fused to the activation domain (AD) and DNA binding domain (DBD) of
831 the yeast Gal4 protein. Interaction between them reconstitutes a functional Gal4 and allows
832 cells to grow on the dropout media (right column), while all strains are able to grow on the rich
833 media (left column). "Vector" indicates a Gal4 DBD or Gal4 AD only plasmid without a bait or
834 target fused to it and is included as a control. Two independent transformants are shown.
835

836 **Fig. 7 Summary of Pho4 protein divergence and its contribution to the Pho2-**
837 **dependence variation in ScPho4 and CgPho4.** Left (1): CgPho4 DBD recognizes the
838 same DNA sequences as ScPho4 DBD, but binds the consensus motif with a 3-4 fold higher
839 affinity. While consistent with the enhanced activity model, further studies found no evidence for
840 the enhanced activity to be a major factor for the reduced Pho2-dependence. (2) Two regions in
841 the IDR of CgPho4 (red asterisks) can enhance the activity of the main activation domain (AD)
842 but are not sufficient on their own to activate gene expression. These two Activation Enhancing
843 Domains (AEDs) directly contribute to CgPho4's reduced Pho2-dependence. Right (3): An IDR
844 (green asterisk) adjacent to the DBD in ScPho4, named P2ID, can both interact with ScPho2
845 and boost the activity of Pho4 and also repress its activity when Pho2 is absent, providing
846 support for the autoinhibition model.
847
848 **Fig. S1 ScPho4 and CgPho4 structural prediction shows both proteins are largely**
849 **disordered outside the DBD.** (A, B) Secondary structure predictions for ScPho4 and CgPho4
850 using the PSIPRED 4.0 server. (C, D) AlphaFold2 predicted monomer 3D structure for ScPho4
851 and CgPho4. Color represents model confidence. It has been shown that low and very low
852 confidence regions overlap significantly and are predictive of protein disorder.
853
854 **Fig. S2 ScPho4 and CgPho4 alignment and annotation.** Protein sequence alignment was
855 generated using ProbCons with 2 passes of consistency transformation and 100 iterative
856 refinements and was manually edited to align the known Pho85 recognition motifs (red lines).
857 The shading behind the amino acid symbols represents the level of conservation based on
858 BLOSUM62 scores. The functional annotation on the top were based on ScPho4. The orange
859 boxes and texts refer to the breakpoints used for constructing chimeric Pho4 constructs (BR4-4
860 only used in Fig. 6).

861

862 **Fig. S3 Genome context Protein Binding Microarray (gcPBM) reveals little divergence in**
863 **sequence preference between ScPho4 and CgPho4.** (A) 36-bp probes were designed with
864 ChIP-identified binding peaks for both ScPho4 and CgPho4 in their respective genome,
865 centered on the consensus E-box motif, CACGTG, along with the flanking nucleotides.
866 Recombinantly purified full length ScPho4 and CgPho4 were hybridized to separate arrays, and
867 the log signal intensities by the two Pho4 proteins for the same oligo were plotted as a
868 scatterplot. A Spearman's rank correlation coefficient and the *P*-value based on a permutation
869 test for the correlation being 0 (no correlation) are provided at the top of the graph. Number of
870 probes is provided in the lower right. (B) Same as A, except the 36-bp probes were ChIP-
871 identified peaks with a non-consensus E-box motifs (mostly with 1-bp-mismatch to the
872 consensus).

873

874 **Fig. S4 Electrophoretic Mobility Shift Assay (EMSA) confirms difference in binding**
875 **affinity to the consensus motif between the two orthologous Pho4 DBDs.** (A) EMSA with
876 ScPho4 DBD and IR700 labeled 17bp consensus DNA. Lanes 2-8 contain 2x diluted protein;
877 lane 1 is a no protein (NP) control. (B) Fraction of bound DNA was calculated from the unbound
878 DNA band normalized to the NP lane. A one-site specific-binding with Hill coefficient model was
879 fit to the data in GraphPad Prism v10. Estimated dissociation constant K_d was shown. (C) and
880 (D) are the same as (A) and (B), but for CgPho4 DBD. Two replicates were performed for each
881 protein and yielded consistent results.

882

883 **Fig. S5 PADDLE prediction for synthetic yeast one-hybrid constructs revealed no new**
884 **regions with activation potential.** PADDLE was run on the full length yeast one-hybrid
885 constructs as was done for the endogenous ScPho4 and CgPho4. Results were plotted using a
886 fixed x axis ranging from 0 to 350 aa even though most constructs were shorter. Like in Fig. 3,

887 the first and last 26 aa had no score due to the design of the prediction algorithm. The greyed
888 region indicates the location of the Gal4 DBD. Its first 22 positions received a Z-score < 0 and
889 hence not visible in the plot. Two constructs had peaks exceeding the Z-score limit of 10. They
890 were plotted with an extended y-axis limit of 0 to 12 (red asterisks). The horizontal dotted lines
891 indicate the thresholds for medium and strong activation potentials, respectively.
892
893 **Fig. S6 Testing Pho2-dependence for the activation potential in the Yeast-1-Hybrid (Y1H)**
894 **assay.** (A) Y1H was performed both with and without *PHO2* (gray and orange bars,
895 respectively). "Host" = no plasmids All names are as specified in Fig. 3 of the main text. Bars
896 represent the means of n>3 replicates; individual data points are shown as dots; lines are the
897 standard deviations. Student's t-tests were used to compare the activity of each construct with
898 vs without Pho2. None of the tested constructs was significant at a 0.05 level after Bonferroni-
899 Holm correction. Full length Gal4 was tested in the same experiment but plotted separately to
900 avoid compressing the scale. (B) Constructs highlighted by the blue vertical line in the left panel
901 contain CgE1 and were tested for the boosting effect without Pho2 by comparing their activity
902 against the corresponding AD alone. Two-sided student's t-tests were performed and the
903 difference in the mean and raw *P*-values were reported.
904
905 **Fig. S7 NLS from ScPho4 and CgPho4 conferred similar levels of nuclear localization.** Six
906 Pho4 constructs with either NLS:Sc or NLS:Cg were quantified for nuclear Pho4 concentration
907 and fraction of Pho4 in the nucleus, in either *PHO2* or *pho2*Δ host background. Endogenous or
908 chimeric Pho4 protein levels were quantified by the C-terminal mNeon tag. Nucleus was labeled
909 by DAPI. Nuclear concentration was estimated using the average gray intensity inside the
910 nucleus. Nuclear fraction was calculated by dividing the Integrated density inside the nucleus by
911 the total integrated density in the whole cell. A total of 10 cells were randomly selected from the
912 bright field image per construct in each background. Individual estimates are shown as dots and

913 a boxplot, where the box represents the interquartile range, the middle line represents the
914 median and the whiskers 1.5 times the interquartile range. ANOVA test found no significant
915 differences for either the identity of the NLS ("Cg" or "Sc") or Host (*PHO2* or *pho2*Δ) factors (F-
916 test raw P = 0.41 and 0.15, respectively).
917
918 **Fig. S8 Evolutionary patterns of AED and P2ID length among Pho4 orthologs.** Time-
919 calibrated species tree on the left is based on Shen et al. 2018 (PMID: 30415838). Pho4 is a
920 single gene family in all budding yeasts, including the 8 species shown. Among these Pho4
921 orthologs, three showed reduced dependence on Pho2 based on He et al. 2017 (PMID:
922 28485712). Middle: P2ID length is calculated based on the multiple sequence alignment
923 constructed using ProbCons (Supplementary data 1). Right: activation potential is predicted
924 using PADDLE. Medium (Z-score between 4 and 6) and strong (Z-score > 6) activation potential
925 regions are shown as light and dark blue stripes. The bHLH DBD is shown as a green box for
926 reference.
927
928 **Fig. S9 Example gating strategy for flow cytometry data.** (A) All events were plotted on
929 FSC.H and SSC.H. A rectangular gate is used to exclude non-cell events. (B) Events within the
930 first gate ("cell") were plotted on FSC.H and FSC.W. Singlets (single cell event as opposed to
931 doublets or multilets) were selected by excluding events with a higher FSC.W given the same
932 FSC.H.

933 **TABLES**

934 **Table 1. Breakpoints for chimeric Pho4 and individual regions tested in this study**

Origin	Region Name	Start	End	Length	Used in figure(s)
CgPho4	R1	1	44	44	Fig. 1,4,5,6,S2,S4
CgPho4	AD	45	112	68	Fig. 1,4,5,6,S2,S4
CgPho4	NLS	113	282	170	Fig. 1,4,5,6,S2,S4
CgPho4	P2ID	283	458	176	Fig. 1,4,5,6,S2,S4
CgPho4	DBD	459	533	75	Fig. 1,4,5,6,S2,S4
ScPho4	R1	1	42	42	Fig. 1,4,5,6,S2,S4
ScPho4	AD	43	99	57	Fig. 1,4,5,6,S2,S4
ScPho4	NLS	100	176	77	Fig. 1,4,5,6,S2,S4
ScPho4	P2ID	177	242	66	Fig. 1,4,5,6,S2,S4
ScPho4	DBD	243	312	70	Fig. 1,4,5,6,S2,S4
CgPho4	P2ID first half	283	327	45	Fig. 6, S2
CgPho4	P2ID second half	328	458	131	Fig. 6, S2
ScPho4	P2ID first half	177	204	28	Fig. 6, S2
ScPho4	P2ID second half	205	242	38	Fig. 6, S2
CgPho4	E1	2	44	43	Fig. 3,S5
CgPho4	AD	45	126	82	Fig. 3,S5
CgPho4	E2	252	463	212	Fig. 3,S5
CgPho4	AD.1	45	75	31	Fig. 3,S5
CgPho4	AD.2	76	126	51	Fig. 3,S5
CgPho4	E2_9aa	252	288	37	Fig. 3,S5
ScPho4	E1	2	42	41	Fig. 3,S5
ScPho4	AD	43	101	59	Fig. 3,S5
ScPho4	AD_9aa	70	101	32	Fig. 3,S5
ScPho4	E2	156	250	95	Fig. 3,S5
ScPho2	AD	404	559	156	Fig. 3
ScPho4	Alternative NLS	100	156	57	Fig. 5
CgPho4	Alternative P2ID	253	463	211	Fig. 5
ScPho4	ΔbHLH	2	250	249	Fig. 6 Y2H
CgPho4	ΔbHLH	2	463	462	Fig.6 Y2H
ScPho2	Pho4int	112	407	296	Fig. 6 Y2H
CgPho4	purified DBD	452	533	82	Fig. 2,S4
ScPho4	purified DBD	236	312	77	Fig. 2,S4

935

936 **Table 2. Plasmids used in this study**

ID	Plasmid	Derived	Source	Comment
pH246	pGAD-C3		PMID:8978031, gift from Fassler lab	Yeast two-hybrid
pH247	pGBD-C3		PMID:8978031, gift from Fassler lab	Yeast one-hybrid
pH343	pRS306-GAL1pr-mCherry	pRS306	this study	Yeast one-hybrid
pH346	pRS314-GAL4pr-Gal4	pRS314	this study	Yeast one-hybrid
pH349	pGBD-Gal4DBD-ScE1	pH247	this study	Yeast one-hybrid
pH350	pGBD-Gal4DBD-CgE1	pH247	this study	Yeast one-hybrid
pH351	pGBD-Gal4DBD-ScE1-ScAD	pH247	this study	Yeast one-hybrid
pH352	pGBD-Gal4DBD-CgE1-CgAD	pH247	this study	Yeast one-hybrid
pH362	pGBD-Gal4DBD-CgAD	pH247	this study	Yeast one-hybrid
pH363	pGBD-Gal4DBD-ScAD	pH247	this study	Yeast one-hybrid
pH377	pGBD-Gal4DBD-CgE2	pH247	this study	Yeast one-hybrid
pH378	pGBD-Gal4DBD-ScE2	pH247	this study	Yeast one-hybrid
pH379	pGBD-Gal4DBD-ScAD.9aa	pH247	this study	Yeast one-hybrid
pH380	pGBD-Gal4DBD-CgAD.2	pH247	this study	Yeast one-hybrid
pH386	pGBD-Gal4DBD-CgE2.9aa	pH247	this study	Yeast one-hybrid
pH387	pGBD-Gal4DBD-CgAD.1	pH247	this study	Yeast one-hybrid
pH388	pGBD-Gal4DBD-CgE1-ScAD.9aa	pH247	this study	Yeast one-hybrid
pH389	pGBD-Gal4DBD-ScAD.9aa-CgE1	pH247	this study	Yeast one-hybrid
pH390	pGBD-Gal4DBD-ScAD.9aa-CgE2.9aa	pH247	this study	Yeast one-hybrid
pH391	pGBD-Gal4DBD-CgE2.9aa-ScAD.9aa	pH247	this study	Yeast one-hybrid
pH434	pGBD-Gal4DBD-ScPho2AD	pH247	this study	Yeast one-hybrid
pH394	pGAD-ScPho4ΔDBD	pH246	this study	Yeast two-hybrid
pH395	pGBD-Pho2Pho4int	pH247	this study	Yeast two-hybrid
pH396	pGAD-CgPho4ΔDBD	pH246	this study	Yeast two-hybrid
pH050	pET-11a-ScPho4 DBD-6xHis	pET11a	this study	BLI
pH051	pET-11a-CgPho4 DBD-6xHis	pET11a	this study	BLI
bH404	pET-11a-GST-CgPho4	pET11a	this study	PBM
pH073	bRA89, Cas9, CEN/ARS, HygR		PMID: 28405019 gift from Malkova lab	CRISPR
pH173	ScPHO4pr, mNeon, CEN/ARS, LEU2	pRS315	this study	Chimera backbone
pH188	Cg(1-533)	pH173	this study	Chimera
pH194	Sc(1-312)	pH173	this study	Chimera

pH209	Cg(1-112) Sc(100-176) Cg(283-533)	pH173	this study	Chimera
pH210	Cg(1-458) Sc(243-312)	pH173	this study	Chimera
pH211	Cg(1-282) Sc(177-242) Cg(459-533)	pH173	this study	Chimera
pH212	Sc(1-42) Cg(45-533)	pH173	this study	Chimera
pH213	Cg(1-44) Sc(43-99) Cg(113-533)	pH173	this study	Chimera
pH215	Sc(1-99) Cg(113-533)	pH173	this study	Chimera
pH216	Sc(1-42) Cg(45-458) Sc(243-312)	pH173	this study	Chimera
pH217	Cg(1-44) Sc(43-176) Cg(283-533)	pH173	this study	Chimera
pH218	Cg(1-112) Sc(100-242) Cg(459-533)	pH173	this study	Chimera
pH219	Cg(1-282) Sc(177-312)	pH173	this study	Chimera
pH220	Sc(1-42) Cg(45-112) Sc(100-176) Cg(283-533)	pH173	this study	Chimera
pH221	Cg(1-44) Sc(43-99) Cg(113-282) Sc(177-312)	pH173	this study	Chimera
pH222	Sc(1-42) Cg(45-282) Sc(177-312)	pH173	this study	Chimera
pH223	Cg(1-44) Sc(43-99) Cg(113-282) Sc(177-242) Cg(459-533)	pH173	this study	Chimera
pH224	Sc(1-42) Cg(45-112) Sc(100-176) Cg(283-458) Sc(243-312)	pH173	this study	Chimera
pH227	Sc(1-176) Cg(283-533)	pH173	this study	Chimera
pH229	Sc(1-176) Cg(283-458) Sc(243-312)	pH173	this study	Chimera
pH230	Sc(1-99) Cg(113-282) Sc(177-312)	pH173	this study	Chimera
pH231	Cg(1-44) Sc(43-312)	pH173	this study	Chimera
pH232	Sc(1-42) Cg(45-112) Sc(100-312)	pH173	this study	Chimera
pH233	Sc(1-242) Cg(459-533)	pH173	this study	Chimera
pH234	Cg(1-112) Sc(100-312)	pH173	this study	Chimera
pH235	Sc(1-99) Cg(113-458) Sc(243-312)	pH173	this study	Chimera
pH236	Cg(1-76) Sc(71-99) Cg(113-282) Sc(177-312)	pH173	this study	Chimera
pH237	Cg(1-44) Sc(43-70) Cg(77-282) Sc(177-312)	pH173	this study	Chimera
pH239	Sc(1-99) Cg(113-282) Sc(177-242) Cg(459-533)	pH173	this study	Chimera
pH240	Cg(1-112) Sc(100-176) Cg(283-458) Sc(243-312)	pH173	this study	Chimera

pH241	Cg(1-44) Sc(43-242) Cg(459-533)	pH173	this study	Chimera
pH250	Cg(1-44) Sc(43-176) Cg(283-458) Sc(243-312)	pH173	this study	Chimera
pH251	Sc(1-42) Cg(45-282) Sc(177-242) Cg(459-533)	pH173	this study	Chimera
pH252	Sc(1-42) Cg(45-112) Sc(100-242) Cg(459-533)	pH173	this study	Chimera
pH253	Cg(1-44) Sc(43-99) Cg(113-458) Sc(243-312)	pH173	this study	Chimera
pH254	Cg(1-44) Sc(43-176) Cg(283-327) Sc(205-242) Cg(459-533)	pH173	this study	Chimera
pH255	Cg(1-44) Sc(43-204) Cg(328-533)	pH173	this study	Chimera
pH256	Cg(1-44) Sc(43-220) Cg(377-533)	pH173	this study	Chimera
pH257	Cg(1-282) Sc(177-204) Cg(328-533)	pH173	this study	Chimera
pH258	Cg(1-327) Sc(205-242) Cg(459-533)	pH173	this study	Chimera
pH265	Sc(1-204) Cg(328-533)	pH173	this study	Chimera
pH266	Sc(1-153) Cg(250-533)	pH173	this study	Chimera
pH276	Sc(1-42) Cg(45-282) Sc(177-204) Cg(328-533)	pH173	this study	Chimera
pH277	Cg(1-44) Sc(43-99) Cg(113-282) Sc(177-204) Cg(328-533)	pH173	this study	Chimera
pH278	Sc(1-204) Cg(328-458) Sc(243-312)	pH173	this study	Chimera
pH279	Cg(1-44) Sc(43-204) Cg(328-458) Sc(243-312)	pH173	this study	Chimera
pH294	Cg(1-470) Sc(251-312)	pH173	this study	Chimera
pH301	Sc(1-247) Cg(464-533)	pH173	this study	Chimera
pH326	Sc(1-156) Cg(253-533)	pH173	this study	Chimera
pH327	Cg(1-252) Sc(157-312)	pH173	this study	Chimera
pH328	Sc(1-42) Cg(45-252) Sc(157-312)	pH173	this study	Chimera
pH329	Cg(1-44) Sc(43-247) Cg(464-533)	pH173	this study	Chimera
pH330	Cg(1-44) Sc(43-156) Cg(253-463) Sc(251-312)	pH173	this study	Chimera
pH331	Cg(1-44) Sc(43-156) Cg(253-533)	pH173	this study	Chimera
pH332	Sc(1-42) Cg(45-463) Sc(251-312)	pH173	this study	Chimera
pH334	Sc(1-42) Cg(45-252) Sc(157-247) Cg(464-533)	pH173	this study	Chimera

pH436	Cg(1-112) Sc(100-176) Cg(283-458) Sc(177-312)	pH173	this study	Chimera
pH438	Sc(1-176) Cg(283-458) Sc(177-312)	pH173	this study	Chimera
pH440	Cg(1-44) Sc(43-99) Cg(113-458) Sc(177-312)	pH173	this study	Chimera
pH441	Sc(1-42) Cg(45-458) Sc(177-312)	pH173	this study	Chimera
pH442	Sc(1-99) Cg(113-458) Sc(177-312)	pH173	this study	Chimera

Table 3. Yeast strains

ID	Genotype	Source
yH373	pho4::NAT; pho2::HIS3; pho80::TRP1; pho5::mCherry	this study
yH555	pho4::NAT; his3::pRS303-HIS3; pho80::TRP1; pho5::mCherry	this study
yH549	gal4delta gal80delta LYS2::GAL1-HIS3 GAL2-ADE2 MET2::GAL7-lacZ	PMID: 8978031, gift from Fassler lab
yH696	gal4delta gal80delta LYS2::GAL1-HIS3; GAL2-ADE2; MET2::GAL7-lacZ; ura3::pRS306-*GAL1pr*-mCherry-URA3	this study
yH156	pho80::TRP1	PMID: 28485712
various	chimeric Pho4 plasmids transformed into yH373 and yH555	this study
various	yeast two-hybrid plasmids transformed into yH696	this study
various	yeast one-hybrid plasmids transformed into yH696	this study

REFERENCES

Anand R, Memisoglu G, Haber J. 2017. Cas9-mediated gene editing in Saccharomyces cerevisiae.

Ang Y-S, Rivas RN, Ribeiro AJS, Srivas R, Rivera J, Stone NR, Pratt K, Mohamed TMA, Fu J-D, Spencer CI, Tippens ND, Li M, Narasimha A, Radzinsky E, Moon-Grady AJ, Yu H, Pruitt BL, Snyder MP, Srivastava D. 2016. Disease Model of GATA4 Mutation Reveals Transcription Factor Cooperativity in Human Cardiogenesis. *Cell* **167**:1734-1749.e22. doi:10.1016/j.cell.2016.11.033

Avsec Ž, Weilert M, Shrikumar A, Krueger S, Alexandari A, Dalal K, Fropf R, McAnany C, Gagneur J, Kundaje A, Zeitlinger J. 2021. Base-resolution models of transcription-factor binding reveal soft motif syntax. *Nat Genet*. doi:10.1038/s41588-021-00782-6

Baker CR, Booth LN, Sorrells TR, Johnson AD. 2012. Protein modularity, cooperative binding, and hybrid regulatory states underlie transcriptional network diversification. *Cell* **151**:80–95. doi:10.1016/j.cell.2012.08.018

Barbaric S, Münsterkötter M, Goding C, Hörz W. 1998. Cooperative Pho2-Pho4 Interactions at the PHO5 Promoter Are Critical for Binding of Pho4 to UASp1 and for Efficient Transactivation by Pho4 at UASp2. *Mol Cell Biol* **18**:2629–2639.

Berger MF, Badis G, Gehrke AR, Talukder S, Philippakis AA, Peña-Castillo L, Alleyne TM, Mnaimneh S, Botvinnik OB, Chan ET, Khalid F, Zhang W, Newburger D, Jaeger SA,

960 Morris QD, Bulyk ML, Hughes TR. 2008. Variation in homeodomain DNA binding
961 revealed by high-resolution analysis of sequence preferences. *Cell* **133**:1266–1276.
962 doi:10.1016/j.cell.2008.05.024
963 Berger MF, Bulyk ML. 2009. Universal protein-binding microarrays for the comprehensive
964 characterization of the DNA-binding specificities of transcription factors. *Nat Protoc*
965 **4**:393–411. doi:10.1038/nprot.2008.195
966 Berger MF, Bulyk ML. 2006. Protein Binding Microarrays (PBMs) for Rapid, High-Throughput
967 Characterization of the Sequence Specificities of DNA Binding ProteinsGene Mapping,
968 Discovery, and Expression. New Jersey: Humana Press. pp. 245–260. doi:10.1385/1-
969 59745-097-9:245
970 Bhoite LT, Allen JM, Garcia E, Thomas LR, Gregory ID, Voth WP, Whelihan K, Rolfes RJ,
971 Stillman DJ. 2002. Mutations in the Pho2 (Bas2) Transcription Factor That Differentially
972 Affect Activation with Its Partner Proteins Bas1, Pho4, and Swi5. *J Biol Chem*
973 **277**:37612–37618.
974 Britton CS, Sorrells TR, Johnson AD. 2020. Protein-coding changes preceded cis-regulatory
975 gains in a newly evolved transcription circuit. *Science* **367**:96–100.
976 doi:10.1126/science.aax5217
977 Brodsky S, Jana T, Mittelman K, Chapal M, Kumar DK, Carmi M, Barkai N. 2020. Intrinsically
978 Disordered Regions Direct Transcription Factor In Vivo Binding Specificity. *Mol Cell*.
979 doi:10.1016/j.molcel.2020.05.032
980 Buchan DWA, Jones DT. 2019. The PSIPRED Protein Analysis Workbench: 20 years on.
981 *Nucleic Acids Res* **47**:W402–W407. doi:10.1093/nar/gkz297
982 Do CB, Mahabhashyam MSP, Brudno M, Batzoglou S. 2005. ProbCons: Probabilistic
983 consistency-based multiple sequence alignment. *Genome Res* **15**:330–340.
984 doi:10.1101/gr.2821705
985 Donovan BT, Chen H, Eek P, Meng Z, Jipa C, Tan S, Bai L, Poirier MG. 2023. Basic helix-loop-
986 helix pioneer factors interact with the histone octamer to invade nucleosomes and
987 generate nucleosome-depleted regions. *Mol Cell* **83**:1251-1263.e6.
988 doi:10.1016/j.molcel.2023.03.006
989 Ferrie JJ, Karr JP, Graham TGW, Dailey GM, Zhang G, Tjian R, Darzacq X. 2023. p300 is an
990 obligate integrator of combinatorial transcription factor inputs. *Mol Cell* S1097-
991 2765(23)01023–7. doi:10.1016/j.molcel.2023.12.004
992 Ferrie JJ, Karr JP, Tjian R, Darzacq X. 2022. “Structure”-function relationships in eukaryotic
993 transcription factors: The role of intrinsically disordered regions in gene regulation. *Mol*
994 *Cell* **0**. doi:10.1016/j.molcel.2022.09.021
995 Gietz RD, Schiestl RH. 2007. High-efficiency yeast transformation using the LiAc/SS carrier
996 DNA/PEG method. *Nat Protoc* **2**:31–34. doi:10.1038/nprot.2007.13
997 Gomes-Vieira AL, Wideman JG, Paes-Vieira L, Gomes SL, Richards TA, Meyer-Fernandes JR.
998 2018. Evolutionary conservation of a core fungal phosphate homeostasis pathway
999 coupled to development in Blastocladiella emersonii. *Fungal Genet Biol* **115**:20–32.
1000 doi:10.1016/j.fgb.2018.04.004
1001 Gordân R, Shen N, Dror I, Zhou T, Horton J, Rohs R, Bulyk ML. 2013. Genomic regions flanking
1002 E-box binding sites influence DNA binding specificity of bHLH transcription factors
1003 through DNA shape. *Cell Rep* **3**:1093–1104. doi:10.1016/j.celrep.2013.03.014
1004 Guss KA, Nelson CE, Hudson A, Kraus ME, Carroll SB. 2001. Control of a Genetic Regulatory
1005 Network by a Selector Gene. *Science* **292**:1164–1167. doi:10.1126/science.1058312
1006 He BZ, Zhou X, O’Shea EK. 2017. Evolution of reduced co-activator dependence led to target
1007 expansion of a starvation response pathway. *eLife* **6**. doi:10.7554/eLife.25157
1008 Hirst K, Fisher F, McAndrew PC, Goding CR. 1994a. The transcription factor, the Cdk, its cyclin
1009 and their regulator: directing the transcriptional response to a nutritional signal. *EMBO J*
1010 **13**:5410–5420.

1011 Hirst K, Fisher F, McAndrew PC, Goding CR. 1994b. The transcription factor, the Cdk, its cyclin
1012 and their regulator: directing the transcriptional response to a nutritional signal. *EMBO J*
1013 **13**:5410–5420.
1014 Hsu IS, Strome B, Lash E, Robbins N, Cowen LE, Moses AM. 2021. A functionally divergent
1015 intrinsically disordered region underlying the conservation of stochastic signaling. *PLOS*
1016 *Genet* **17**:e1009629. doi:10.1371/journal.pgen.1009629
1017 James P, Halladay J, Craig EA. 1996. Genomic libraries and a host strain designed for highly
1018 efficient two-hybrid selection in yeast. *Genetics* **144**:1425–1436.
1019 doi:10.1093/genetics/144.4.1425
1020 Jayaraman PS, Hirst K, Goding CR. 1994. The activation domain of a basic helix-loop-helix
1021 protein is masked by repressor interaction with domains distinct from that required for
1022 transcription regulation. *EMBO J* **13**:2192–2199.
1023 Ji D, Shao C, Yu J, Hou Y, Gao X, Wu Y, Wang L, Chen P. 2024. FOXA1 forms biomolecular
1024 condensates that unpack condensed chromatin to function as a pioneer factor. *Mol Cell*
1025 **84**:244-260.e7. doi:10.1016/j.molcel.2023.11.020
1026 Kaffman A, Herskowitz I, Tjian R, O'Shea EK. 1994. Phosphorylation of the transcription factor
1027 PHO4 by a cyclin-CDK complex, PHO80-PHO85. *Science* **263**:1153–1156.
1028 Kerwin CL, Wykoff DD. 2009. Candida glabrata PHO4 Is Necessary and Sufficient for Pho2-
1029 Independent Transcription of Phosphate Starvation Genes. *Genetics* **182**:471–479.
1030 doi:10.1534/genetics.109.101063
1031 Komeili A, O'Shea EK. 1999a. Roles of Phosphorylation Sites in Regulating Activity of the
1032 Transcription Factor Pho4. *Science* **284**:977–980. doi:10.1126/science.284.5416.977
1033 Komeili A, O'Shea EK. 1999b. Roles of Phosphorylation Sites in Regulating Activity of the
1034 Transcription Factor Pho4. *Science* **284**:977–980. doi:10.1126/science.284.5416.977
1035 Krois AS, Dyson HJ, Wright PE. 2018. Long-range regulation of p53 DNA binding by its
1036 intrinsically disordered N-terminal transactivation domain. *Proc Natl Acad Sci*
1037 **115**:E11302–E11310. doi:10.1073/pnas.1814051115
1038 Lam FH, Steger DJ, O'Shea EK. 2008. Chromatin decouples promoter threshold from dynamic
1039 range. *Nature* **453**:246–250.
1040 Liu J, Perumal NB, Oldfield CJ, Su EW, Uversky VN, Dunker AK. 2006. Intrinsic Disorder in
1041 Transcription Factors. *Biochemistry* **45**:6873–6888. doi:10.1021/bi0602718
1042 Lynch VJ, May G, Wagner GP. 2011. Regulatory evolution through divergence of a
1043 phosphoswitch in the transcription factor CEBPB. *Nature* **480**:383–386.
1044 doi:10.1038/nature10595
1045 Maerkl SJ, Quake SR. 2007. A Systems Approach to Measuring the Binding Energy
1046 Landscapes of Transcription Factors. *Science* **315**:233–237.
1047 doi:10.1126/science.1131007
1048 Magbanua JP, Ogawa N, Harashima S, Oshima Y. 1997. The transcriptional activators of the
1049 PHO regulon, Pho4p and Pho2p, interact directly with each other and with components
1050 of the basal transcription machinery in Saccharomyces cerevisiae. *J Biochem (Tokyo)*
1051 **121**:1182–1189.
1052 McAndrew PC, Svaren J, Martin SR, Hörz W, Goding CR. 1998. Requirements for Chromatin
1053 Modulation and Transcription Activation by the Pho4 Acidic Activation Domain. *Mol Cell*
1054 *Biol* **18**:5818–5827. doi:10.1128/MCB.18.10.5818
1055 Merhej J, Thiebaut A, Blugeon C, Pouch J, Ali Chaouche MEA, Camadro J-M, Le Crom S,
1056 Lelandais G, Devaux F. 2016. A Network of Paralogous Stress Response Transcription
1057 Factors in the Human Pathogen Candida glabrata. *Front Microbiol* **7**.
1058 doi:10.3389/fmicb.2016.00645
1059 Mindel V, Brodsky S, Cohen A, Manadre W, Jonas F, Carmi M, Barkai N. 2023. Intrinsically
1060 disordered regions of the Msn2 transcription factor encode multiple functions using

1061 interwoven sequence grammars. *Nucleic Acids Res* gkad1191.
1062 doi:10.1093/nar/gkad1191
1063 Ming Yip H, Cheng S, Olson EJ, Crone M, Maerkl SJ. 2023. Perfect adaptation achieved by
1064 transport limitations governs the inorganic phosphate response in S. cerevisiae. *Proc*
1065 *Natl Acad Sci U S A* **120**:e2212151120. doi:10.1073/pnas.2212151120
1066 Nocedal I, Johnson AD. 2015. How Transcription Networks Evolve and Produce Biological
1067 Novelty. *Cold Spring Harb Symp Quant Biol* **80**:265–274.
1068 doi:10.1101/sqb.2015.80.027557
1069 Nourani A, Utley RT, Allard S, Côté J. 2004. Recruitment of the NuA4 complex poises the PHO5
1070 promoter for chromatin remodeling and activation. *EMBO J* **23**:2597–2607.
1071 doi:10.1038/sj.emboj.7600230
1072 Ogawa N, Oshima Y. 1990. Functional domains of a positive regulatory protein, PHO4, for
1073 transcriptional control of the phosphatase regulon in Saccharomyces cerevisiae. *Mol*
1074 *Cell Biol* **10**:2224–2236. doi:10.1128/mcb.10.5.2224-2236.1990
1075 O'Neill EM, Kaffman A, Jolly ER, O'Shea EK. 1996. Regulation of PHO4 nuclear localization by
1076 the PHO80-PHO85 cyclin-CDK complex. *Science* **271**:209–212.
1077 Piskacek S, Gregor M, Nemethova M, Grabner M, Kovarik P, Piskacek M. 2007. Nine-amino-
1078 acid transactivation domain: establishment and prediction utilities. *Genomics* **89**:756–
1079 768. doi:10.1016/j.ygeno.2007.02.003
1080 Rives A, Meier J, Sercu T, Goyal S, Lin Z, Liu J, Guo D, Ott M, Zitnick CL, Ma J, Fergus R.
1081 2021. Biological structure and function emerge from scaling unsupervised learning to
1082 250 million protein sequences. *Proc Natl Acad Sci* **118**:e2016239118.
1083 doi:10.1073/pnas.2016239118
1084 Sanborn AL, Yeh BT, Feigerle JT, Hao CV, Townshend RJ, Lieberman Aiden E, Dror RO,
1085 Kornberg RD. 2021. Simple biochemical features underlie transcriptional activation
1086 domain diversity and dynamic, fuzzy binding to Mediator. *eLife* **10**:e68068.
1087 doi:10.7554/eLife.68068
1088 Sánchez-Higueras C, Rastogi C, Voutev R, Bussemaker HJ, Mann RS, Hombría JC-G. 2019. In
1089 vivo Hox binding specificity revealed by systematic changes to a single cis regulatory
1090 module. *Nat Commun* **10**:3597. doi:10.1038/s41467-019-11416-1
1091 Shao D, Creasy CL, Bergman LW. 1996. Interaction of Saccharomyces cerevisiae Pho2 with
1092 Pho4 increases the accessibility of the activation domain of Pho4. *Mol Gen Genet MGG*
1093 **251**:358–364. doi:10.1007/BF02172527
1094 Sherbenou DW, Druker BJ. 2007. Applying the discovery of the Philadelphia chromosome. *J*
1095 *Clin Invest* **117**:2067–2074. doi:10.1172/JCI31988
1096 Shimizu T, Toumoto A, Ihara K, Shimizu M, Kyogoku Y, Ogawa N, Oshima Y, Hakoshima T.
1097 1997. Crystal structure of PHO4 bHLH domain–DNA complex: flanking base recognition.
1098 *EMBO J* **16**:4689–4697. doi:10.1093/emboj/16.15.4689
1099 Slattery M, Riley T, Liu P, Abe N, Gomez-Alcala P, Dror I, Zhou T, Rohs R, Honig B,
1100 Bussemaker HJ, Mann RS. 2011. Cofactor binding evokes latent differences in DNA
1101 binding specificity between Hox proteins. *Cell* **147**:1270–1282.
1102 doi:10.1016/j.cell.2011.10.053
1103 Stewart AJ, Hannenhalli S, Plotkin JB. 2012. Why Transcription Factor Binding Sites Are Ten
1104 Nucleotides Long. *Genetics* **192**:973–985. doi:10.1534/genetics.112.143370
1105 Thomas-Chollier M, Herrmann C, Defrance M, Sand O, Thieffry D, Helden J van. 2011. RSAT
1106 peak-motifs: motif analysis in full-size ChIP-seq datasets. *Nucleic Acids Res* gkr1104.
1107 doi:10.1093/nar/gkr1104
1108 Todeschini A-L, Georges A, Veitia RA. 2014. Transcription factors: specific DNA binding and
1109 specific gene regulation. *Trends Genet* **30**:211–219. doi:10.1016/j.tig.2014.04.002
1110 Traven A, Jelicic B, Sopta M. 2006. Yeast Gal4: a transcriptional paradigm revisited. *EMBO Rep*
1111 **7**:496–499. doi:10.1038/sj.embor.7400679

1112 Troshin PV, Procter JB, Sherstnev A, Barton DL, Madeira F, Barton GJ. 2018. JABAWS 2.2
1113 distributed web services for Bioinformatics: protein disorder, conservation and RNA
1114 secondary structure. *Bioinformatics* **34**:1939–1940. doi:10.1093/bioinformatics/bty045
1115 Tsong AE, Tuch BB, Li H, Johnson AD. 2006. Evolution of alternative transcriptional circuits with
1116 identical logic. *Nature* **443**:415–420. doi:10.1038/nature05099
1117 Tuch BB, Galgoczy DJ, Hernday AD, Li H, Johnson AD. 2008. The evolution of combinatorial
1118 gene regulation in fungi. *PLoS Biol* **6**:e38. doi:10.1371/journal.pbio.0060038
1119 Wapinski I, Pfiffner J, French C, Socha A, Thompson DA, Regev A. 2010. Gene duplication and
1120 the evolution of ribosomal protein gene regulation in yeast. *Proc Natl Acad Sci U S A*
1121 **107**:5505–5510. doi:10.1073/pnas.0911905107
1122 Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2--a
1123 multiple sequence alignment editor and analysis workbench. *Bioinforma Oxf Engl*
1124 **25**:1189–1191. doi:10.1093/bioinformatics/btp033
1125 Weeramange CJ, Fairlamb MS, Singh D, Fenton AW, Swint-Kruse L. 2020. The strengths and
1126 limitations of using biolayer interferometry to monitor equilibrium titrations of
1127 biomolecules. *Protein Sci* pro.3827. doi:10.1002/pro.3827
1128 Zhang Y, Ho TD, Buchler NE, Gordân R. 2021. Competition for DNA binding between
1129 paralogous transcription factors determines their genomic occupancy and regulatory
1130 functions. *Genome Res* **31**:1216–1229. doi:10.1101/gr.275145.120
1131 Zhou X, O'Shea EK. 2011. Integrated approaches reveal determinants of genome-wide binding
1132 and function of the transcription factor Pho4. *Mol Cell* **42**:826–836.
1133 doi:10.1016/j.molcel.2011.05.025
1134
1135

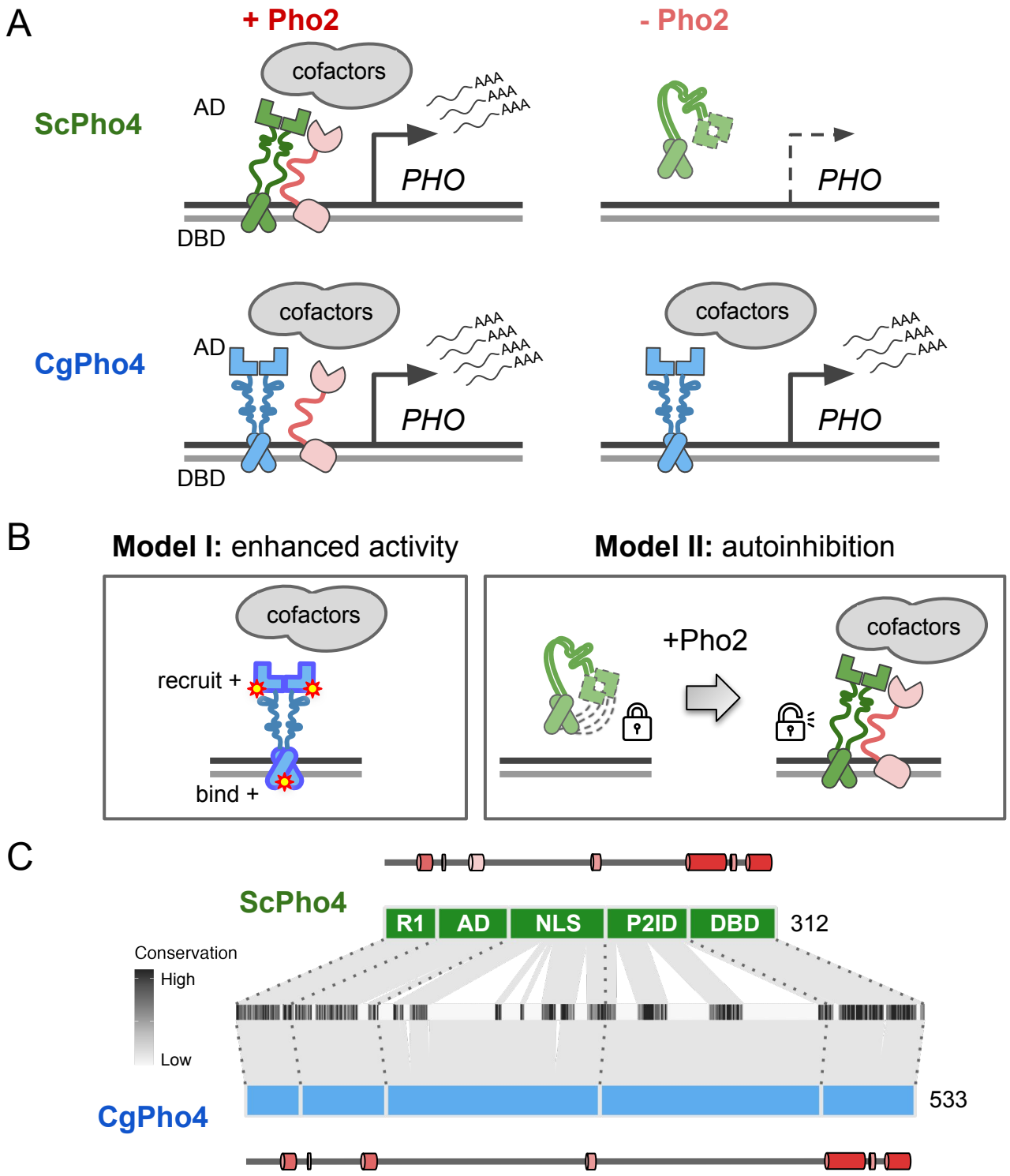


Fig. 1 Different co-TF (Pho2) dependence between orthologous TF (Pho4) in yeast species. (A) Cartoon summary of previous work: *S. cerevisiae* Pho4 (ScPho4) depends on the co-TF, Pho2, for DNA binding and activation (top row), while Pho4 from *C. glabrata* (CgPho4) can bind and activate without the co-TF. See text for references. (B) Two *non-mutually exclusive* models to explain the divergence in Pho2-dependence. (C) The tiles in the middle show the conservation scores at aligned positions of the two Pho4 orthologs; the box diagrams represent ScPho4 and CgPho4, with gray lines connecting matching residues to the alignment tiles. Five regions were delineated based on knowledge about ScPho4: R1 = regulatory, interact with Pho80; AD = activation domain; NLS = nuclear import and export signals; P2ID = Pho2-interaction domain (also includes a region contacting Pho80); DBD = basic Helix-Loop-Helix DNA binding domain. The lines and cylinders above and below the region diagrams show predicted α helices in each protein, with lighter color corresponding to lower prediction confidence. Prediction was done using PSIPRED 4.0.

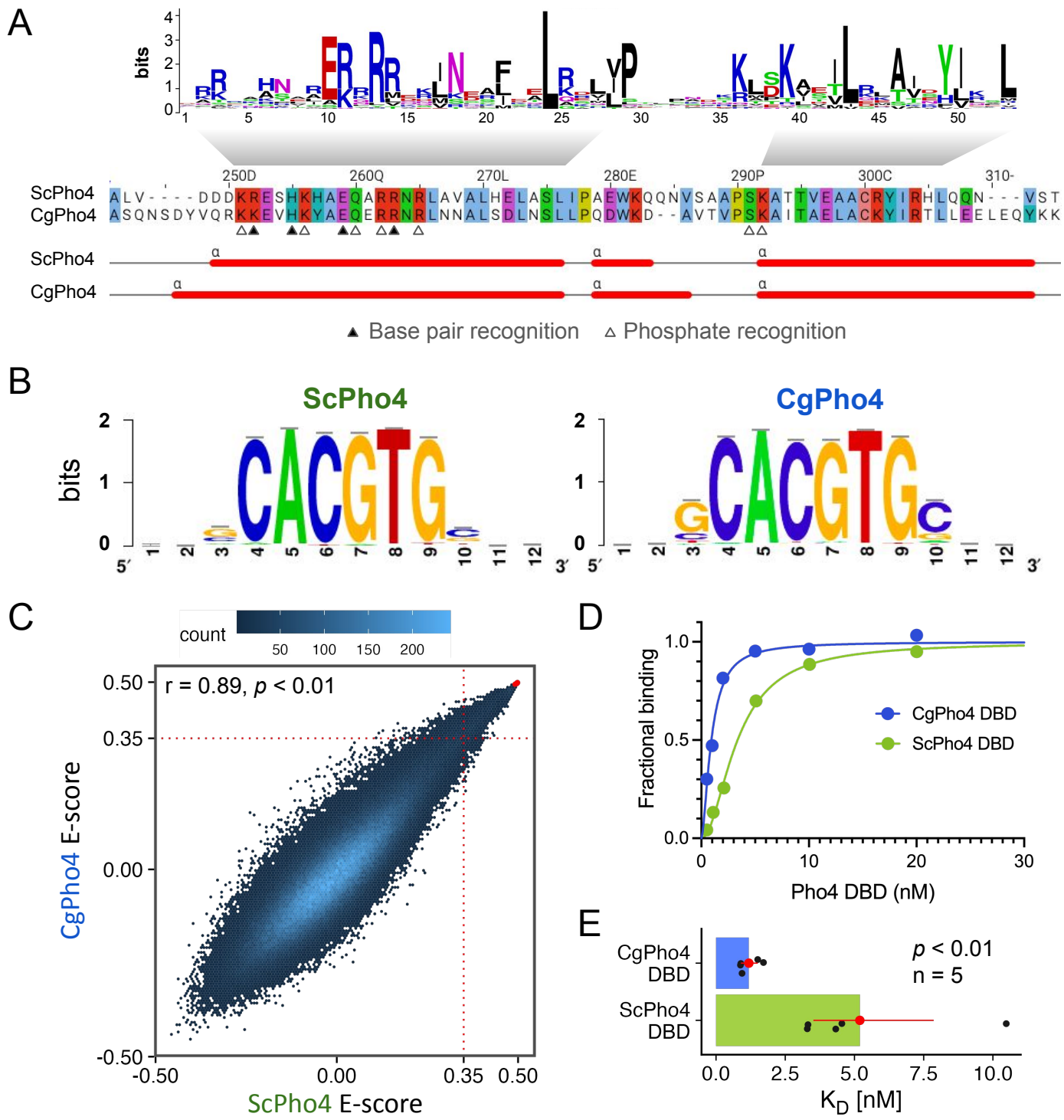


Fig. 2 CgPho4 DBD recognizes the same motif as ScPho4 DBD and has a higher affinity towards the consensus DNA. (A) Alignment of the DBD region between the two Pho4 orthologs. Top: sequence logo for the bHLH domain (PROSITE PS50888). Middle: coordinates above are based on ScPho4. Filled and open triangles indicate residues in ScPho4 DBD making either base or phosphate backbone contacts (Shimizu *et al*., 1997). Bottom: predicted α-helices in both Pho4 DBDs. (B) WebLogo motifs for ScPho4 and CgPho4 are derived from ChIP-seq peaks. (C) Universal Protein Binding Microarray (uPBM) comparing the binding of full length ScPho4 and CgPho4 against all possible 9-mers. A rank-order based E-score is plotted for all 7-mers, where a score greater than 0.35 indicates specific TF-DNA binding (Gordân *et al*., 2013). Red dots in the top right corner represent oligos with the "CACGTG" motif. (D) Binding to the 17bp oligo containing a consensus E-box motif is measured by Biolayer interferometry (BLI). Fractional binding is plotted vs the concentration of purified ScPho4 or CgPho4 DBD. One of the representative binding curves is shown here. (E) Dissociation constants (K_D) are determined from the binding curves and shown as dots (n = 5). The bars and the red dots are the means and the red lines the 95% confidence intervals from bootstrapping. Two-sided Student's t-test was performed on the log transformed K_D. A Wilcoxon rank-sum test on raw K_D similarly yielded $P < 0.01$.

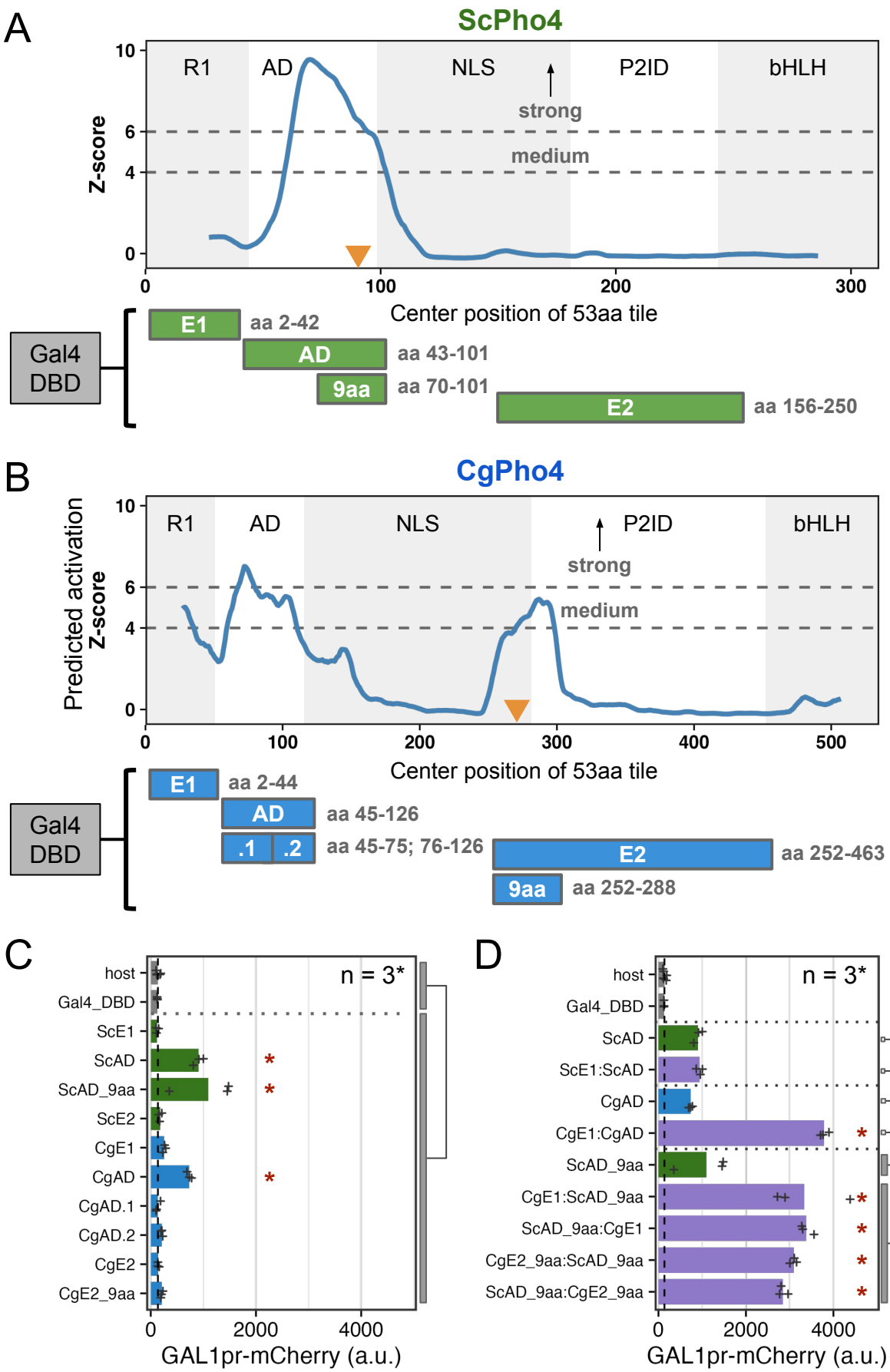

A
ScPho4
R1
AD
NLS
P2ID
bHLH
strong
medium
Z-score
Center position of 53aa tile
Gal4 DBD
E1
aa 2-42
AD
aa 43-101
9aa
aa 70-101
E2
aa 156-250
B
CgPho4
Predicted activation Z-score
Center position of 53aa tile
E1
aa 2-44
AD
aa 45-126
.1
.2
aa 45-75; 76-126
E2
aa 252-463
9aa
aa 252-288
C
n = 3*
host
Gal4_DBD
ScE1
ScAD
ScAD_9aa
ScE2
CgE1
CgAD
CgAD.1
CgAD.2
CgE2
CgE2_9aa
GAL1pr-mCherry (a.u.)
D
n = 3*
host
Gal4_DBD
ScAD
ScE1:ScAD
CgAD
CgE1:CgAD
ScAD_9aa
CgE1:ScAD_9aa
ScAD_9aa:CgE1
CgE2_9aa:ScAD_9aa
ScAD_9aa:CgE2_9aa
GAL1pr-mCherry (a.u.)

Fig. 3 CgPho4 encodes two Activation Enhancer Domains (AEDs) that can enhance the activation potential of the main AD. (A, B) PADDLE prediction of activation potential in ScPho4 and CgPho4. Predictions were made in 53 aa blocks and plotted using the middle of each block as the X coordinate. As a result, the first and last 26 aa positions have no scores. Z-scores > 4 and 6 indicate regions with medium and strong activation potential, respectively. The five regions in each protein are labeled as in Fig. 1. Orange triangles mark the 9aaTAD motif matches. Below the plots are diagrams of the Gal4DBD fusion constructs used to test the activation potential, with their boundaries labeled on the right. C) Activation potential of each region was tested in the context of a Gal4 DBD fusion driving a *GAL1pr*-mCherry reporter. Shown are the mean (bar) and individual data points (n=3, except for “host”, for which n = 6). “host” has the mCherry reporter but no plasmid; “Gal4_DBD” has the reporter and the Gal4 DBD alone. ANOVA was used to compare the Gal4 DBD fusions to the two negative controls combined; an asterisk indicates significant difference at a 0.05 level after Holm-Bonferroni correction. D) CgE1 and CgE2_9aa can enhance the activation capability of ScAD_9aa in an orientation-independent manner. An asterisk indicates significance difference at a 0.05 level by a t-test between the indicated groups after Holm-Bonferroni correction.

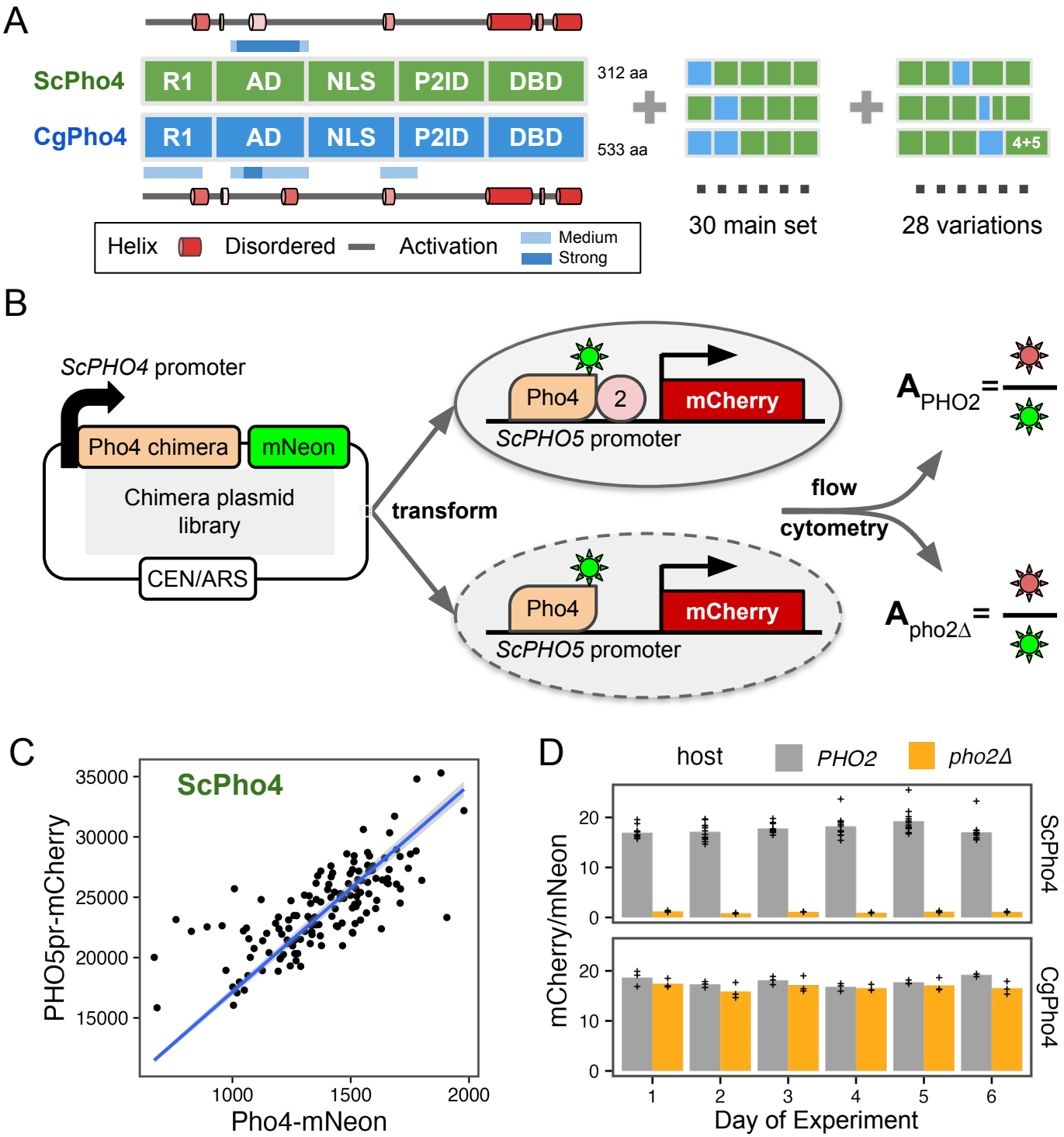


Fig. 4 A dual fluorescence reporter system for accurately quantifying the activity of Pho4 chimeras with and without Pho2. (A) ScPho4 and CgPho4 were divided into five corresponding parts. Names of each region were based on the annotations for ScPho4. Predicted alpha helices were drawn as cartoons outside of the box diagrams. All 2^5 = 32 combinations (2 were the original proteins) were made in the main chimera set, and an additional set of chimeras were constructed to test specific hypotheses. (B) A diagram showing the design of the dual fluorescence reporter system. The chimeric Pho4 fragments were cloned downstream of the endogenous *ScPHO4* promoter and were tagged at the C-terminus with an in-frame mNeon fluorescent protein, followed by the endogenous *ScPHO4* 3'UTR and terminator. The plasmid contains a CEN/ARS element, resulting in stable inheritance and low biological variation. (C) The *PHO5pr-mCherry* levels correlate with the ScPho4-mNeon levels between biological replicates. The linear regression line is shown in blue and a 95% confidence band as a gray ribbon. (D) A_{PHO2} and $A_{pho2\Delta}$ values for ScPho4 and CgPho4 across 6 different days of experiments showing consistent results (n = 3 for all groups except ScPho4 with Pho2, where n = 12)

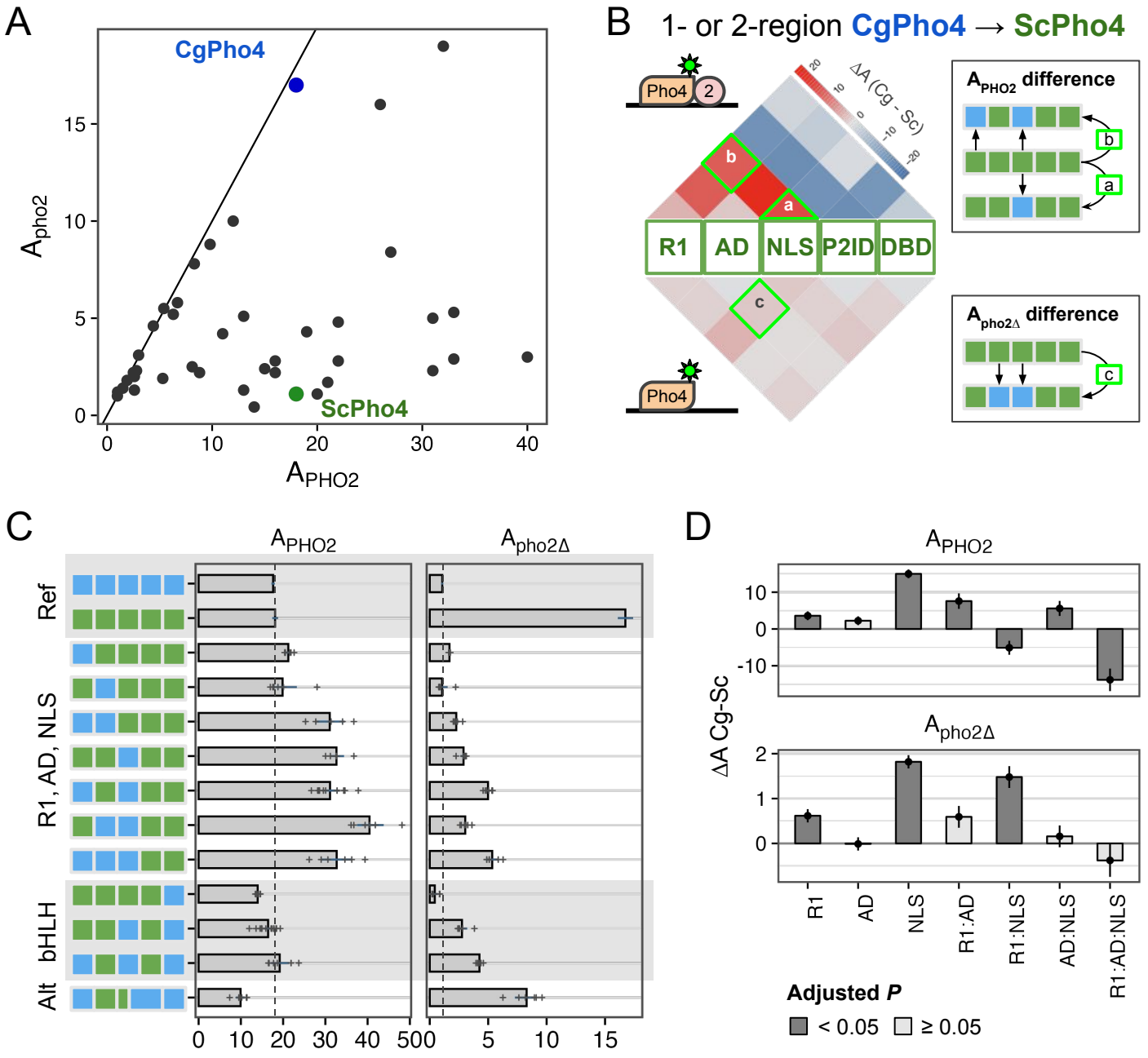


Fig. 5. R1 and NLS of CgPho4 confer stronger activity than their counterparts in ScPho4. (A) Pho4 chimeras (black), endogenous ScPho4 (green) and CgPho4 (blue) were plotted based on their activity with and without Pho2. The solid line indicates equal activity (independent of Pho2). (B) Heatmap showing the difference in activity between the CgPho4 region(s) and the counterpart of ScPho4 measured in the ScPho4 background. The upper half shows the difference with Pho2 and the lower half without Pho2. The triangles (e.g., a) show the difference between single regions while the squares on the top show the differences when swapping two regions (see key to the right). Red color indicates a higher activity of the region from CgPho4 than the counterpart in ScPho4 and blue color indicates decreased activity. (C) A subset of the chimeras were shown with their A_{PHO2} and $A_{pho2\Delta}$ values. Bars represent the means; lines are 95% confidence intervals based on bootstrap; plus signs are individual data points (n>=5), which were not shown for the endogenous ScPho4 and CgPho4 (n>=36). Vertical dotted lines equal the endogenous ScPho4 level. (D) Epistasis analysis of CgPho4 R1, AD and NLS. A linear model was fit to the data with all two and three-way interactions. Shown are estimates of the individual and interaction terms (bar and point) for the CgPho4 region(s). Line ranges indicate the standard errors of the estimates. Dark gray indicates the term is significant at 0.05 level after Holm-Bonferroni correction.

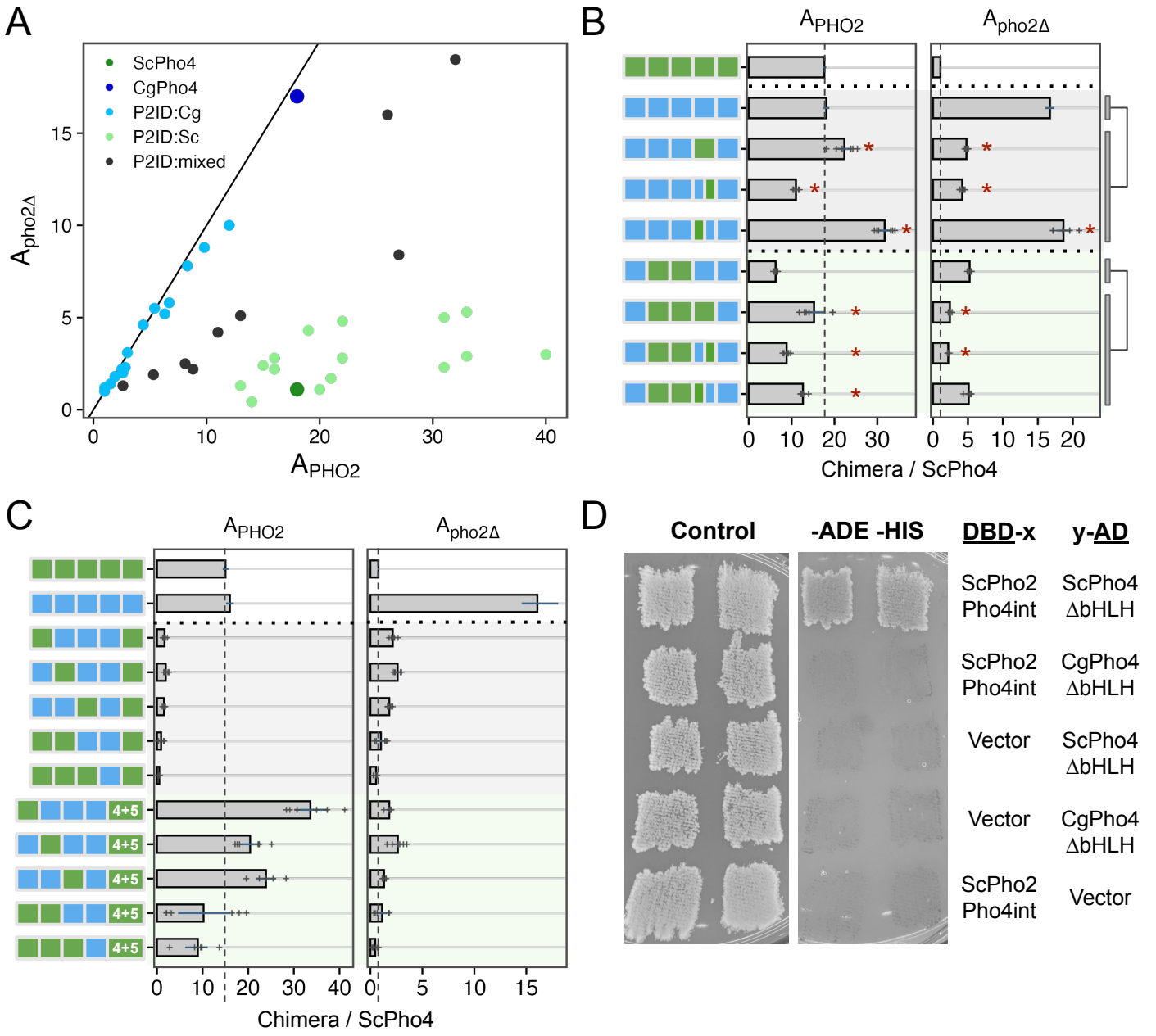


Fig. 6 P2ID in ScPho4 allows for enhanced activation with Pho2 via physical interaction with the co-TF. (A) Scatter plot of all chimeric Pho4 constructs based on their A_{PHO2} and $A_{pho2\Delta}$ values, colored by the identity of their P2ID region. Mixed P2ID uses an additional breakpoint that separates the region into two halves (Fig. S2). (B) Comparing the effects of swapping the whole P2ID from ScPho4 into CgPho4 vs individual halves. Bars represent the means; lines are 95% confidence intervals based on bootstrap; plus signs are individual data points (n>=5), which were not shown for the endogenous ScPho4 and CgPho4 (n>=36). Two sets were tested (rows 2-5 and 6-9). Two-sided Student's t-tests were performed as indicated by the bars and lines to the right. Holm-Bonferroni correction was applied to the raw *P*-values for both sets, separately for A_{PHO2} and $A_{pho2\Delta}$. An asterisk indicates corrected $P < 0.05$. (C) Testing the hypothesis that ScPho4 DBD requires the P2ID:Sc and Pho2 to be fully functional. Plot designs are the same as in (B). Compared with the chimeras with P2ID:Cg + P2ID:Sc, which are all non-functional even with Pho2, adding back P2ID:Sc ("4+5") rescued the activity with Pho2, but not without. (D) Pho4-Pho2 physical interaction was tested with a yeast two hybrid assay. The bait and target regions were fused to the activation domain (AD) and DNA binding domain (DBD) of the yeast Gal4 protein. Interaction between them reconstitutes a functional Gal4 and allows cells to grow on the dropout media (right column), while all strains are able to grow on the rich media (left column). "Vector" indicates a Gal4 DBD or Gal4 AD only plasmid without a bait or target fused to it, and is included as a control. Two independent transformants are shown.

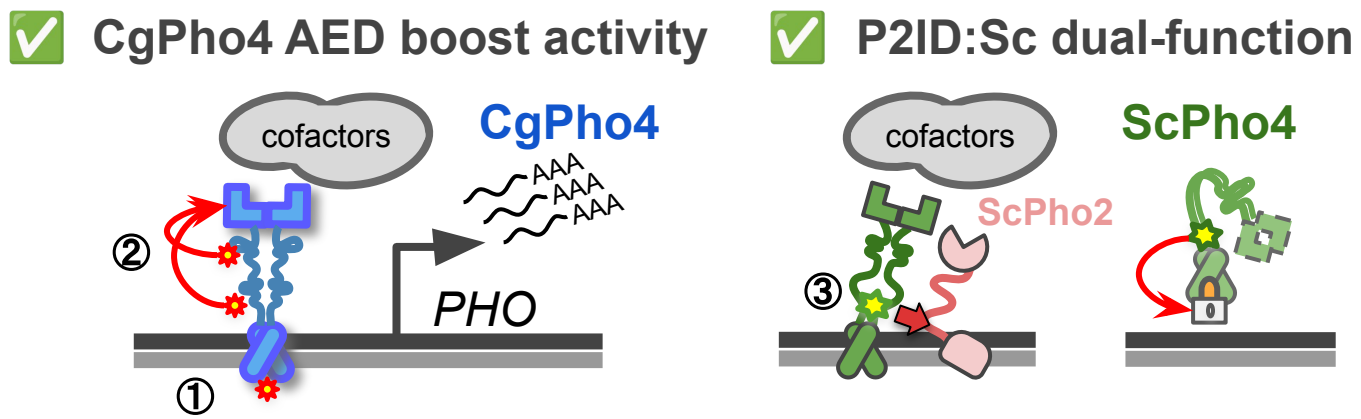


Fig. 7 Summary of Pho4 protein divergence and its contribution to the Pho2-dependence variation in ScPho4 and CgPho4. Left (1): CgPho4 DBD recognizes the same DNA sequences as ScPho4 DBD, but binds the consensus motif with a 3-4 fold higher affinity. While consistent with the enhanced activity model, further studies found no evidence for the enhanced activity to be a major factor for the reduced Pho2-dependence. (2) Two regions in the IDR of CgPho4 (red asterisks) can enhance the activity of the main activation domain (AD) but are not sufficient on their own to activate gene expression. These two Activation Enhancing Domains (AEDs) directly contribute to CgPho4's reduced Pho2-dependence. Right (3): An IDR (green asterisk) adjacent to the DBD in ScPho4, named P2ID, can both interact with ScPho2 and boost the activity of Pho4 and also repress its activity when Pho2 is absent, providing support for the autoinhibition model.

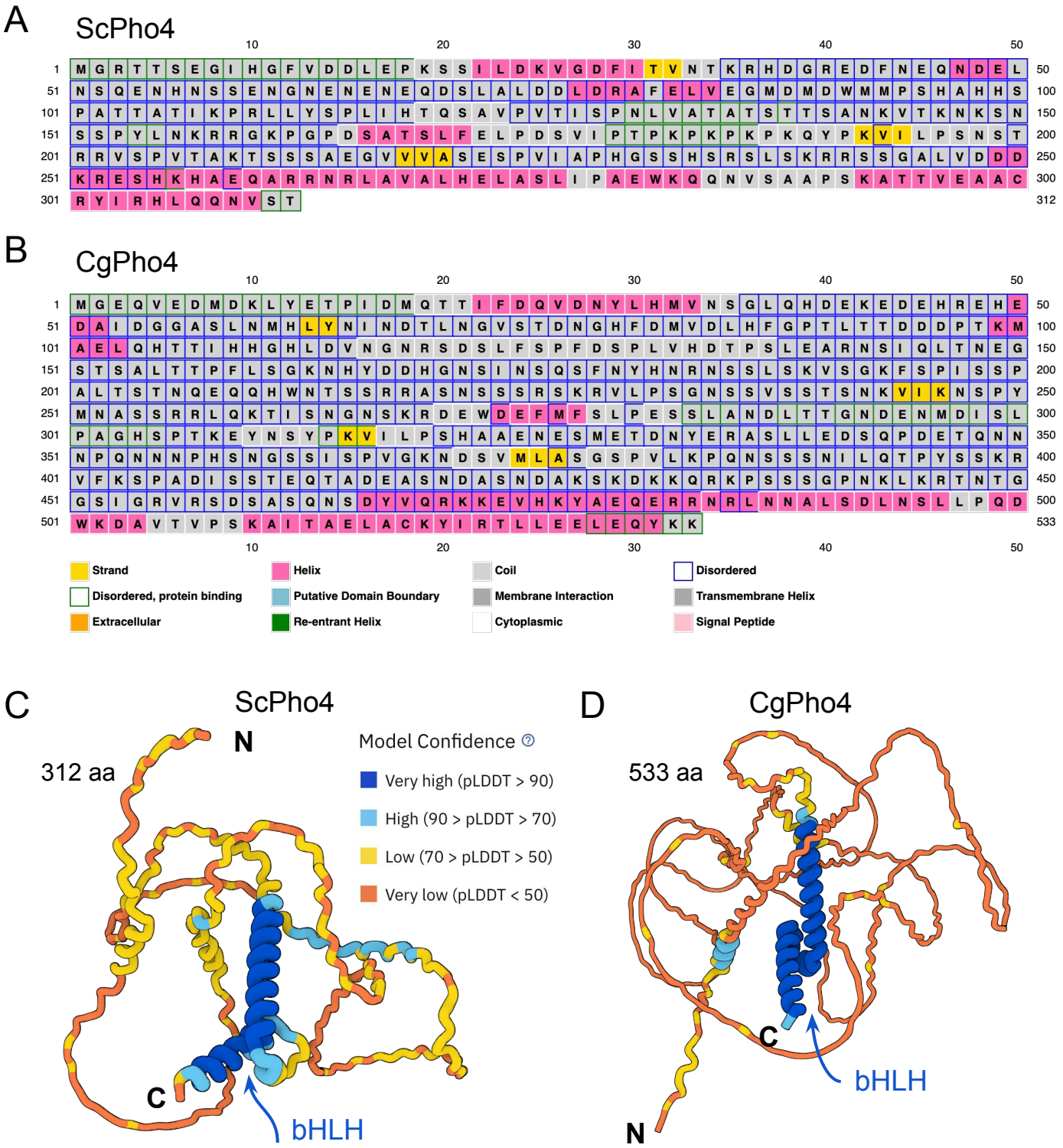


Fig. S1 ScPho4 and CgPho4 structural prediction shows both proteins are largely disordered outside the DBD. (A, B) Secondary structure predictions for ScPho4 and CgPho4 using the PSIPRED 4.0 server. (C, D) AlphaFold2 predicted monomer 3D structure for ScPho4 and CgPho4. Color represents model confidence. It has been shown that low and very low confidence regions overlap significantly and are predictive of protein disorder.

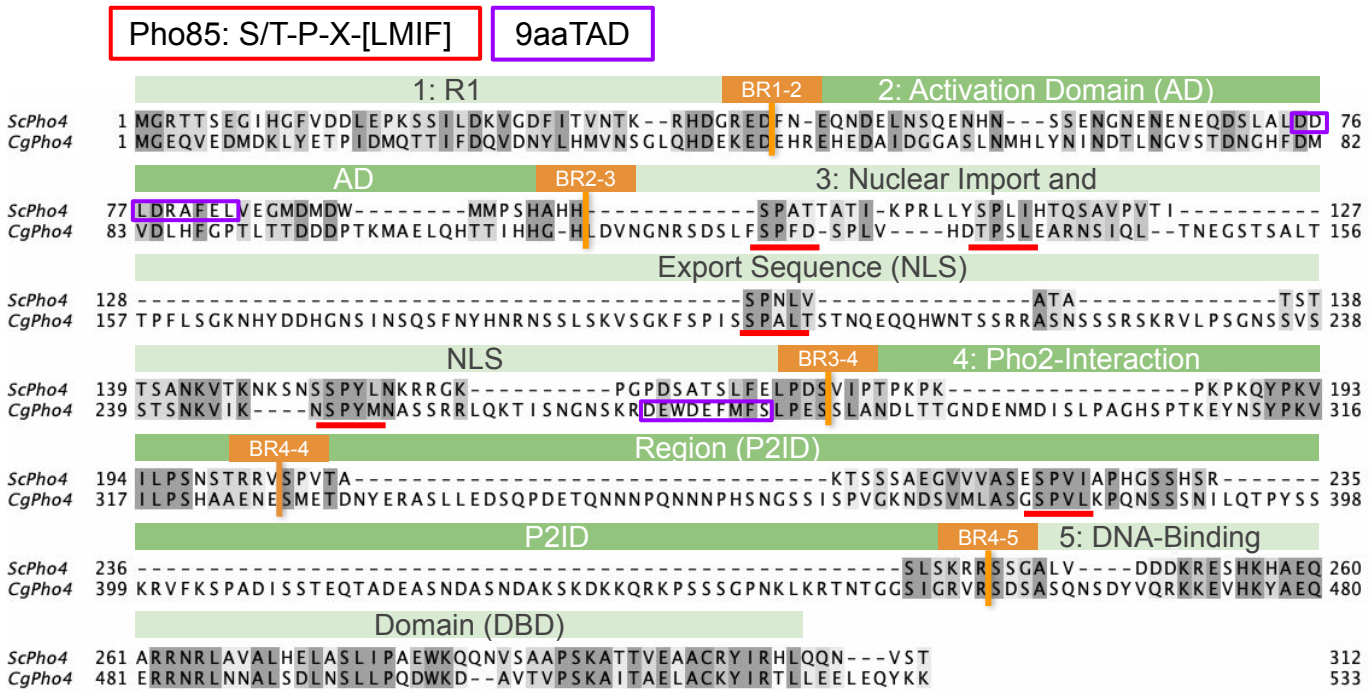


Fig. S2 ScPho4 and CgPho4 alignment and annotation. Protein sequence alignment was generated using ProbCons with 2 passes of consistency transformation and 100 iterative refinement, and was manually edited to align the known Pho85 recognition motifs (red lines). The shading behind the amino acid symbols represent level of conservation based on BLOSUM62 scores. The functional annotation on the top were based on ScPho4. The orange boxes and texts refer to the breakpoints used for constructing chimeric Pho4 constructs (BR4-4 only used in Fig. 6).

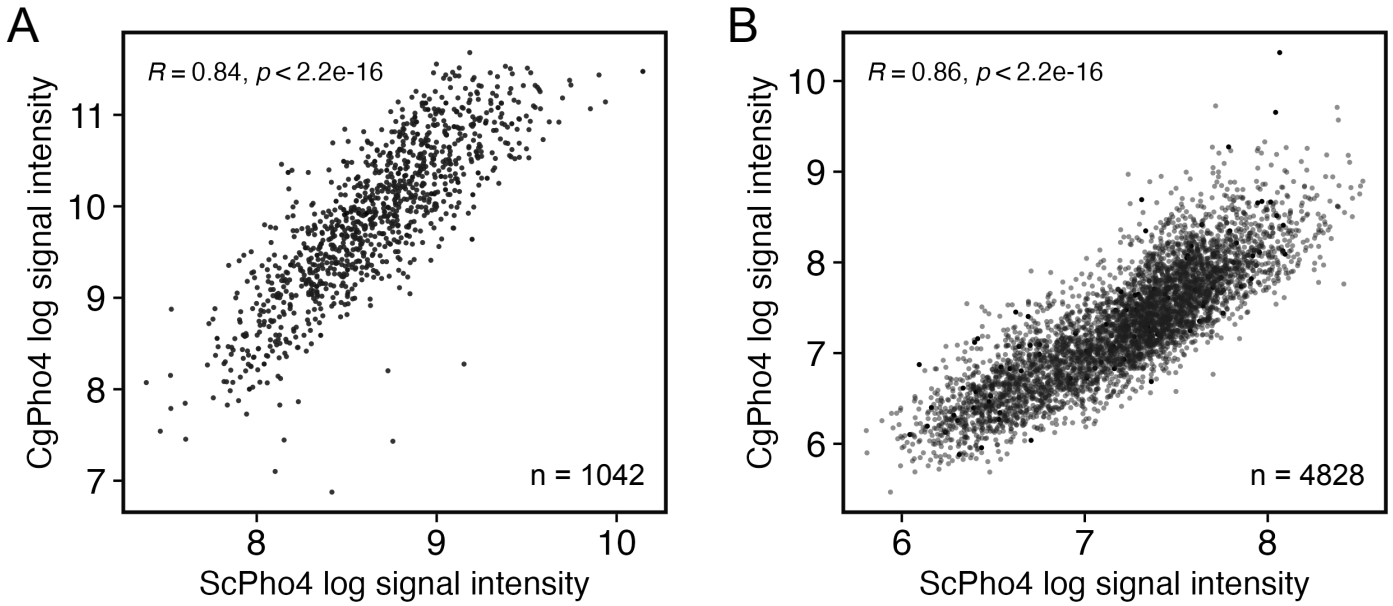


Fig. S3 Genome context Protein Binding Microarray (gcPBM) reveals little divergence in sequence preference between ScPho4 and CgPho4. (A) 36-bp probes were designed with ChIP-identified binding peaks for both ScPho4 and CgPho4 in their respective genome, centered on the consensus E-box motif, CACGTG, along with the flanking nucleotides. Recombinantly purified full length ScPho4 and CgPho4 were hybridized to separate arrays, and the log signal intensities by the two Pho4 proteins for the same oligo were plotted as a scatterplot. A Spearman's rank correlation coefficient and the *P*-value based on a permutation test for the correlation being 0 (no correlation) are provided at the top of the graph. Number of probes is provided in the lower right. (B) Same as A, except the 36-bp probes were ChIP-identified peaks with a non-consensus E-box motifs (mostly with 1-bp-mismatch to the consensus).

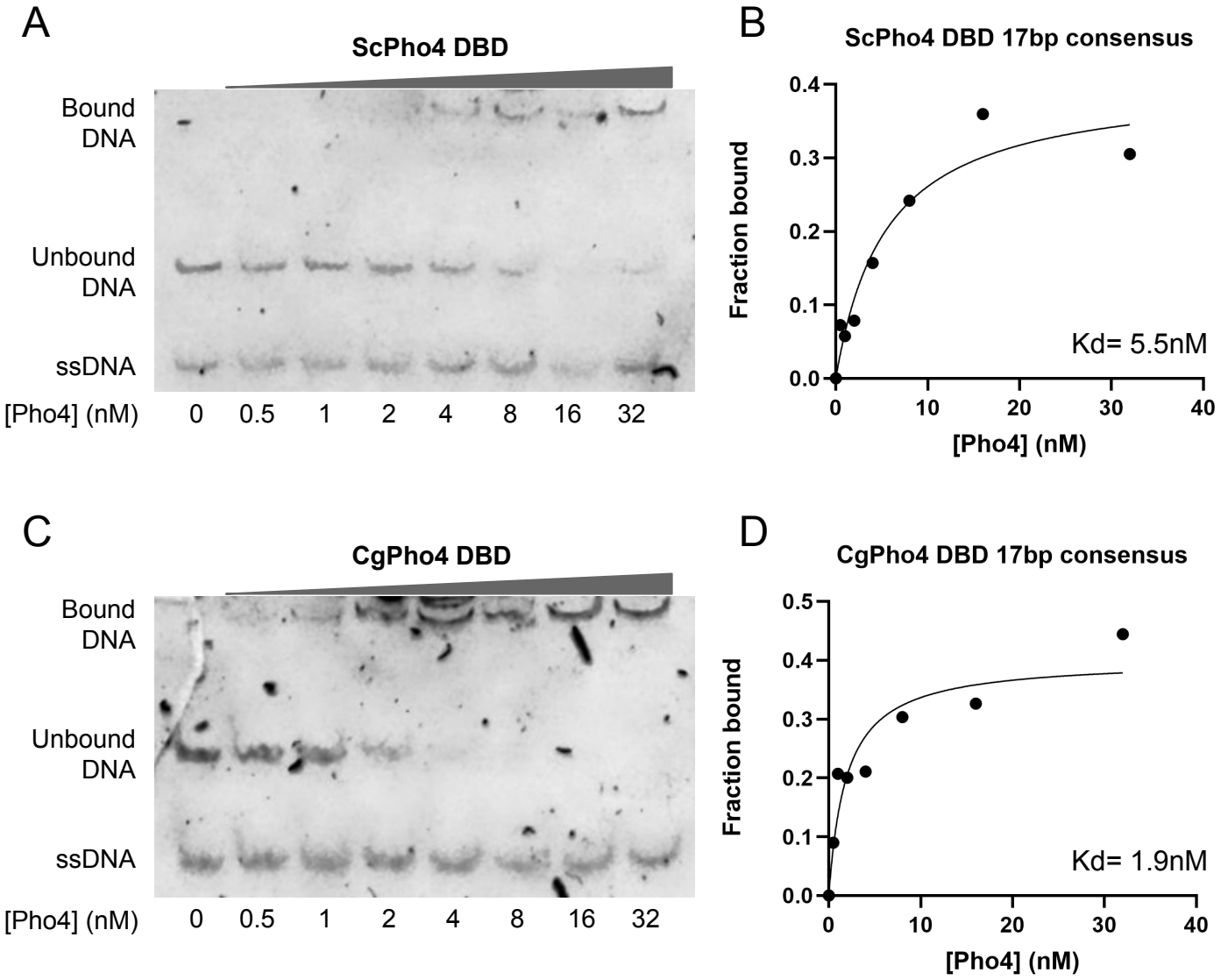


Fig. S4 Electrophoretic Mobility Shift Assay (EMSA) confirms difference in binding affinity to the consensus motif between the two orthologous Pho4 DBDs. (A) EMSA with ScPho4 DBD and IR700 labeled 17bp consensus DNA. Lanes 2-8 contain 2x diluted protein; lane 1 is a no protein (NP) control. (B) the unbound DNA was quantified from (A) and used calculate the fraction bound normalized to the NP lane. A one-site specific-binding with Hill coefficient model was fit to the data in GraphPad Prism v10. Estimated dissociation constant Kd is shown. (C) and (D) are the same as (A) and (B), but for CgPho4 DBD. Two replicates were performed for each protein and yielded consistent results.

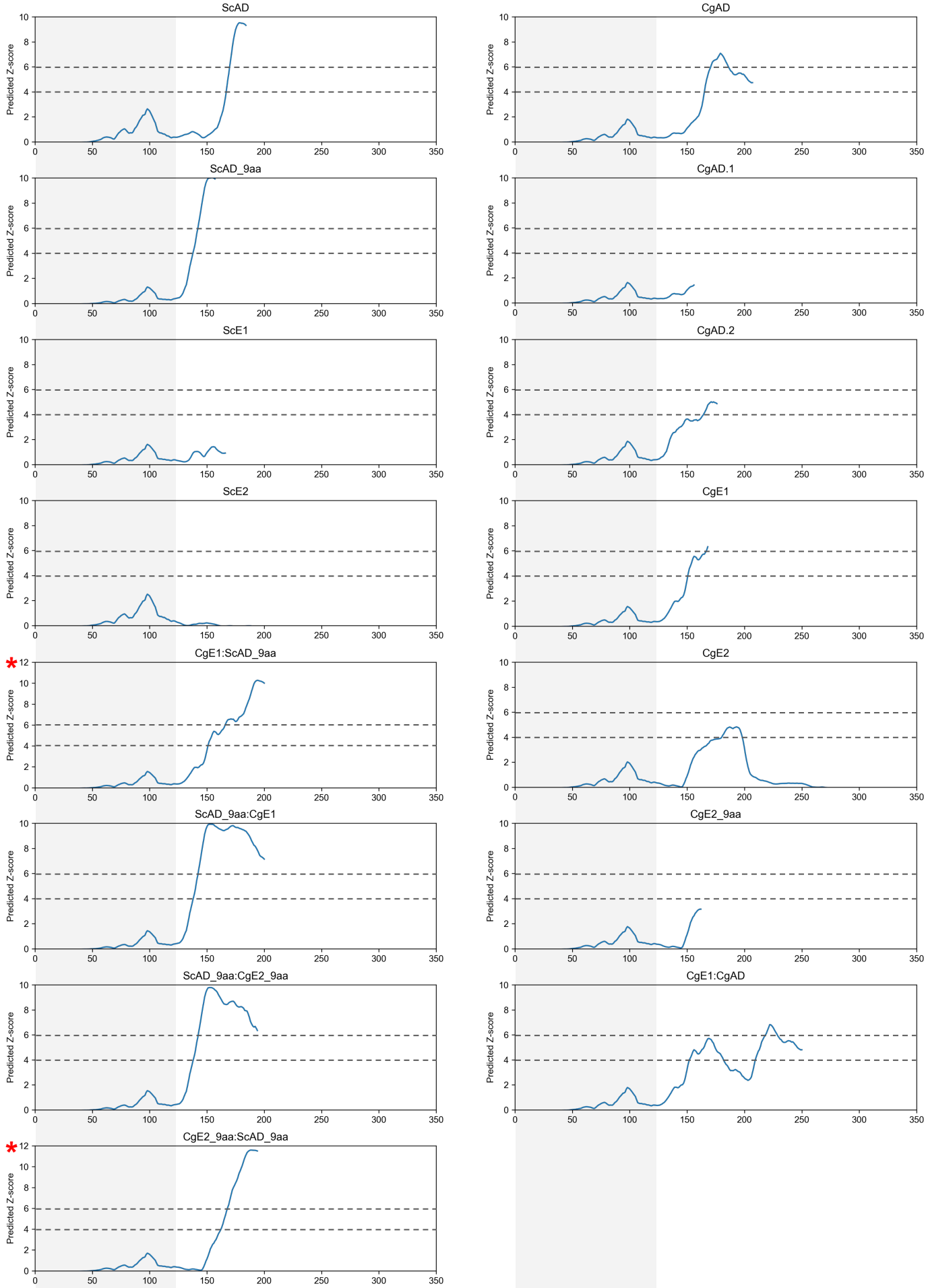
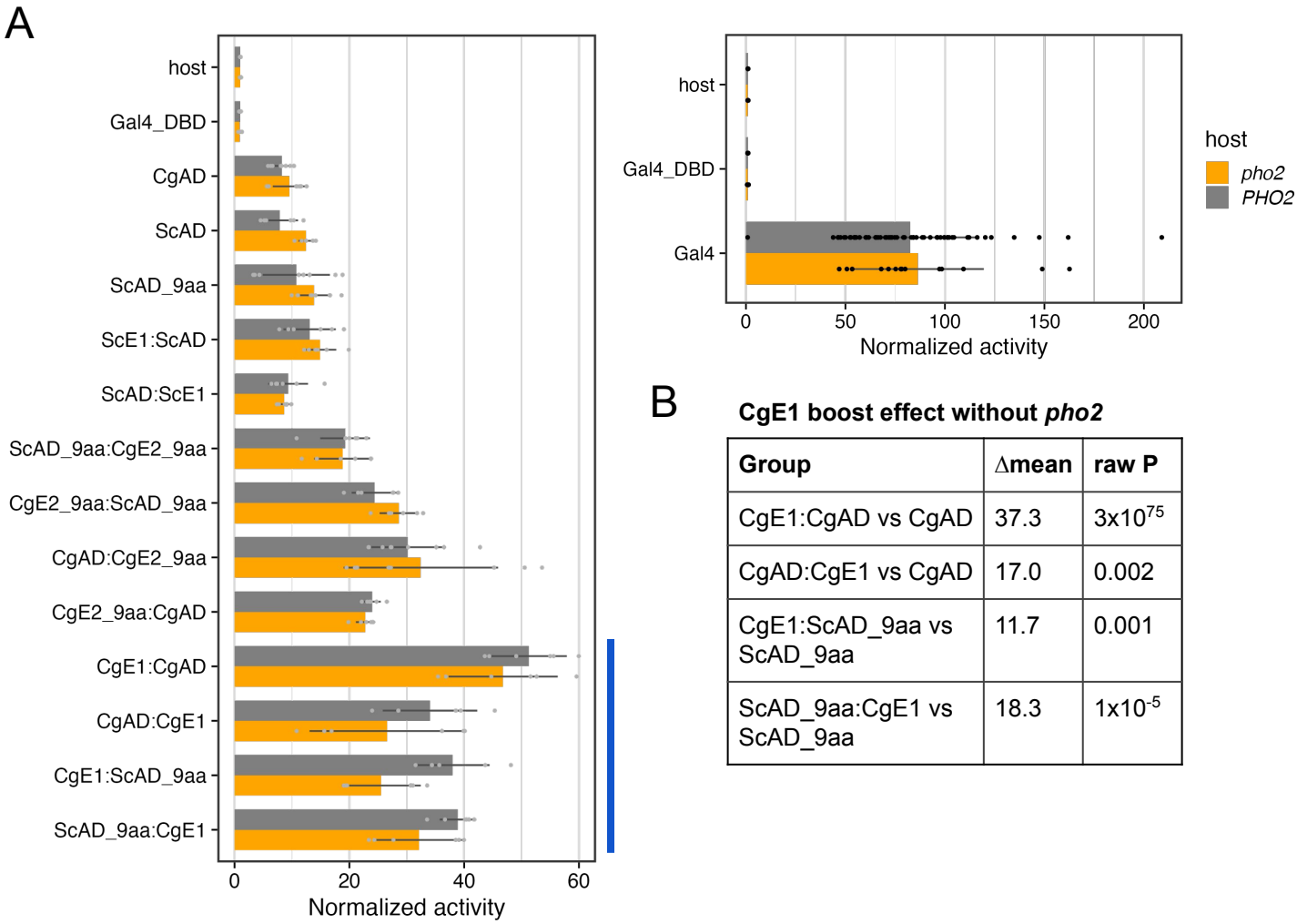

ScAD
CgAD
ScAD_9aa
CgAD.1
ScE1
CgAD.2
ScE2
CgE1
CgE1:ScAD_9aa
CgE2
ScAD_9aa:CgE1
CgE2_9aa
ScAD_9aa:CgE2_9aa
CgE1:CgAD
CgE2_9aa:ScAD_9aa
Predicted Z-score
*
*

Fig. S5 PADDLE prediction for synthetic yeast one-hybrid constructs revealed no new regions with activation potential. PADDLE was run on the full length yeast one-hybrid constructs as was done for the endogenous ScPho4 and CgPho4. Results were plotted using a fixed x axis ranging from 0 to 350 aa even though most constructs were shorter. Like in Fig. 3, the first and last 26 aa had no score due to the design of the prediction algorithm. The greyed region indicates the location of the Gal4 DBD. Its first 22 positions received a Z-score < 0 and hence not visible in the plot. Two constructs had peaks exceeding the Z-score limit of 10. They were plotted with an extended y-axis limit of 0 to 12 (red asterisks). The horizontal dotted lines indicate the thresholds for medium and strong activation potentials, respectively.



B

CgE1 boost effect without *pho2*

Group	Δmean	raw P
CgE1:CgAD vs CgAD	37.3	$3x10^{75}$
CgAD:CgE1 vs CgAD	17.0	0.002
CgE1:ScAD_9aa vs ScAD_9aa	11.7	0.001
ScAD_9aa:CgE1 vs ScAD_9aa	18.3	$1x10^{-5}$

Fig. S6 Testing Pho2-dependence for the activation potential in the Yeast-1-Hybrid (Y1H) assay. (A) Y1H was performed both with and without *PHO2* (gray and orange bars, respectively). "Host" = no plasmids All names are as specified in Fig. 3 of the main text. Bars represent the means of n>3 replicates; individual data points are shown as dots; lines are the standard deviations. Student's t-tests were used to compare the activity of each construct with vs without Pho2. None of the tested constructs was significant at a 0.05 level after Bonferroni-Holm correction. Full length Gal4 was tested in the same experiment and plotted on the right to avoid compressing the scale. (B) Constructs highlighted by the blue vertical line in the left panel contain CgE1 and were tested for the boosting effect without Pho2 by comparing their activity against the corresponding AD alone. Two-sided student's t-tests were performed and the difference in the mean and raw *P*-values were reported.

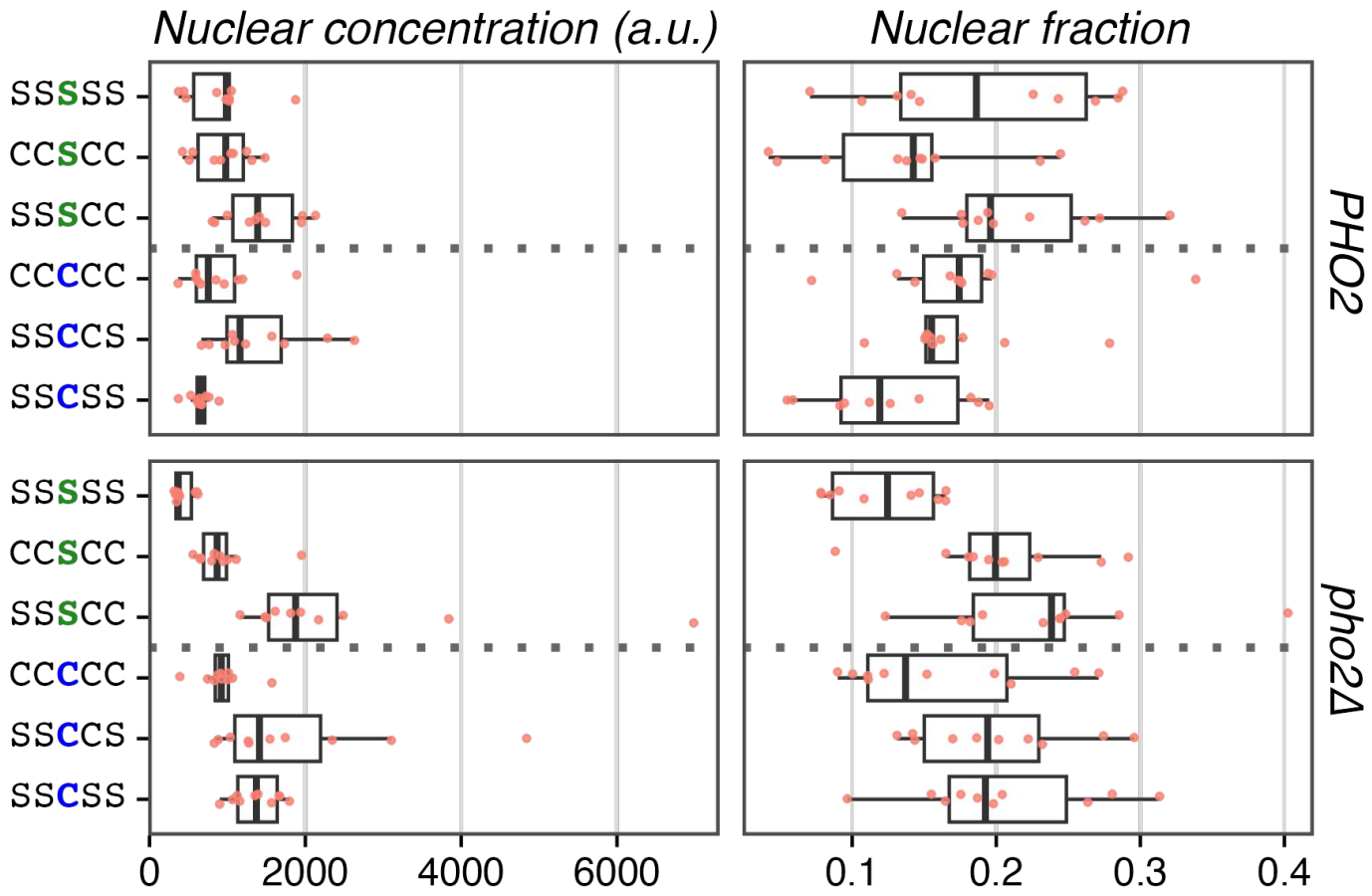


Fig. S7 NLS from ScPho4 and CgPho4 conferred similar levels of nuclear localization. Six Pho4 constructs with either NLS:Sc or NLS:Cg were quantified for nuclear Pho4 concentration and fraction of Pho4 in the nucleus, in either *PHO2* or *pho2*Δ host background. Endogenous or chimeric Pho4 protein levels were quantified by the C-terminal mNeon tag. Nucleus was labeled by DAPI. Nuclear concentration was estimated using the average gray intensity inside the nucleus. Nuclear fraction was calculated by dividing the Integrated density inside the nucleus by the total integrated density in the whole cell. A total of 10 cells were randomly selected from the bright field image per construct in each background. Individual estimates are shown as dots and also as a boxplot, where the box represents the interquartile range, the middle line represents the median and the whiskers 1.5 times the interquartile range. ANOVA test found no significant differences for either the identity of the NLS ("Cg" or "Sc") or Host (*PHO2* or *pho2*Δ) factors (F-test raw P = 0.41 and 0.15, respectively).

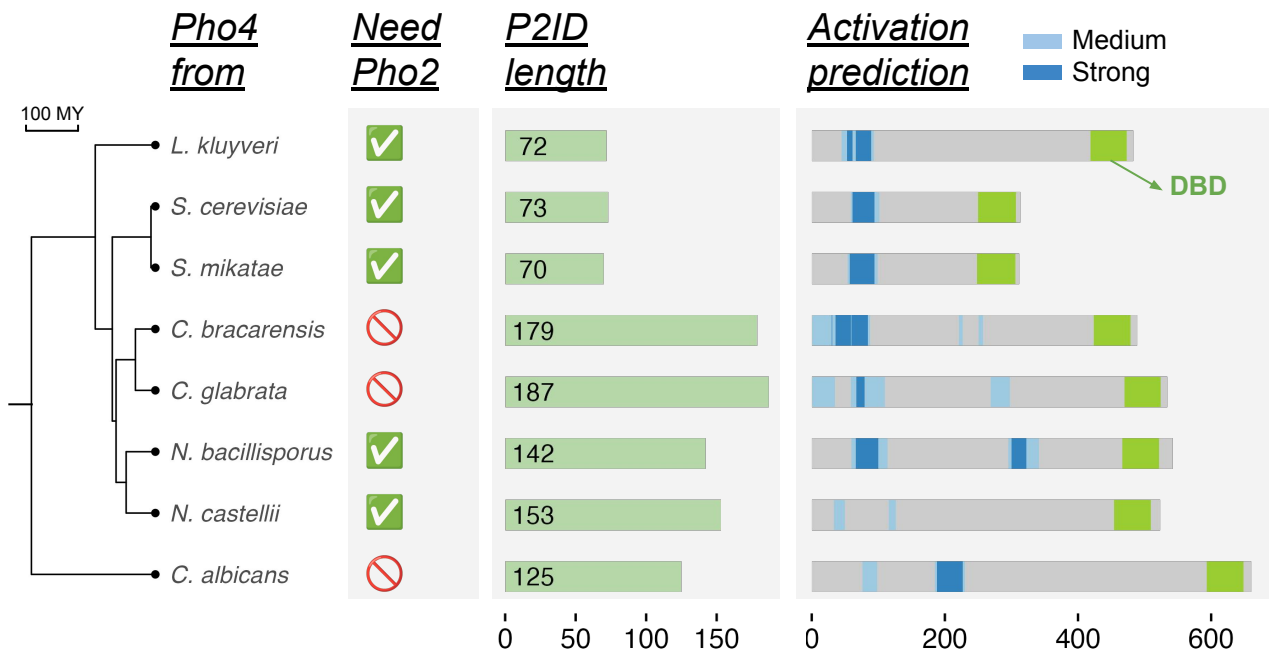


Fig. S8 Evolutionary patterns of AED and P2ID length among Pho4 orthologs. Time-calibrated species tree on the left is based on Shen et al. 2018 (PMID: 30415838). Pho4 is a single gene family in all budding yeasts, including the 8 species shown. Among these Pho4 orthologs, three showed reduced dependence on Pho2 based on He et al. 2017 (PMID: 28485712). Middle: P2ID length is calculated based on the multiple sequence alignment constructed using ProbCons (Supplementary data 1). Right: activation potential is predicted using PADDLE. Medium (Z-score between 4 and 6) and strong (Z-score > 6) activation potential regions are shown as light and dark blue stripes. The bHLH DBD is shown as green box for reference.

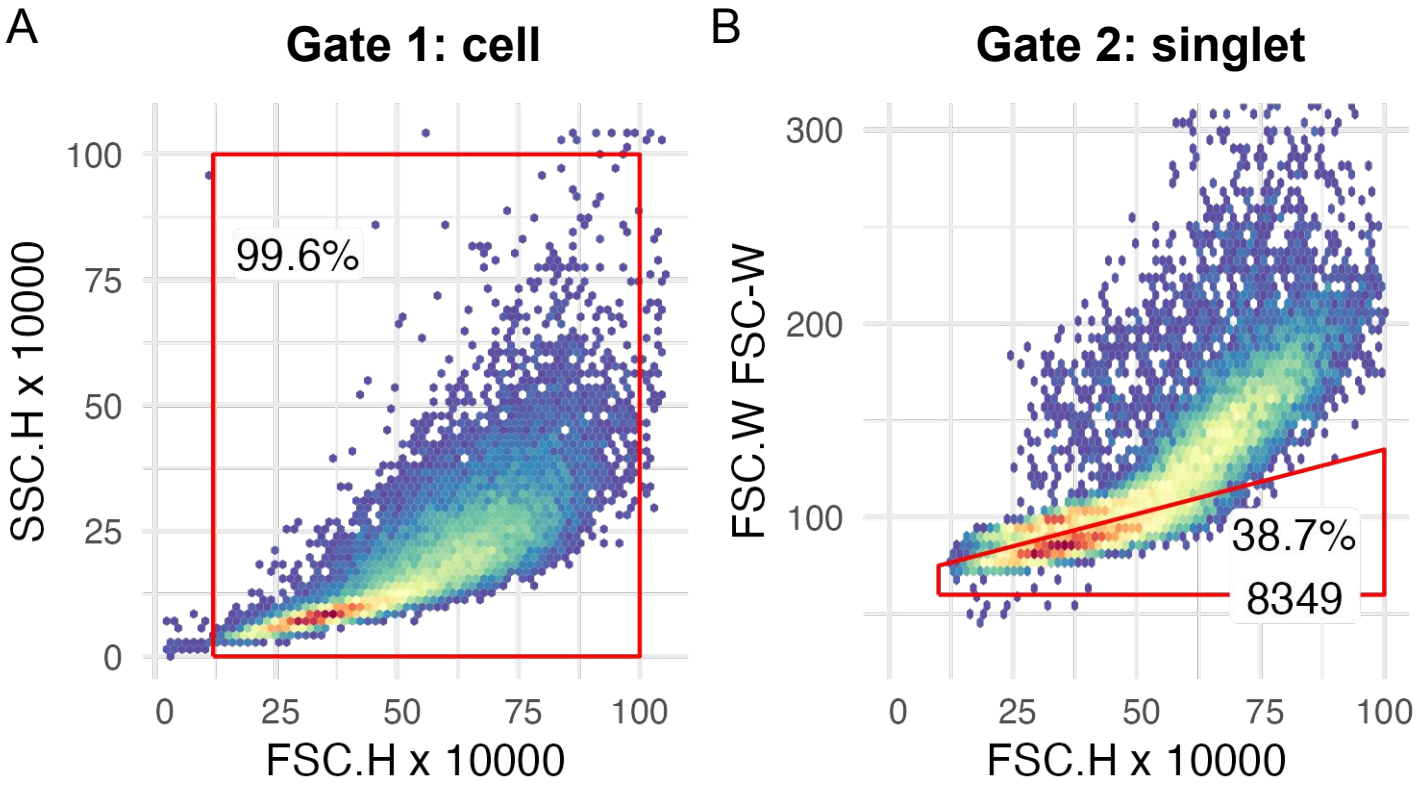


Fig. S9 Example gating strategy for flow cytometry data. (A) All events were plotted on FSC.H and SSC.H. A rectangular gate is used to exclude non-cell events. (B) Events within the first gate ("cell") were plotted on FSC.H and FSC.W. Singlets (single cell event as opposed to doublets or multilets) were selected by excluding events with a higher FSC.W given the same FSC.H.