not necessarily associated with all major impacts [e.g., the Chesapeake impact (41)]. These findings will motivate a search for impact ejecta at other sites to define the geographic footprint of the P-E strewn field, which will ultimately constrain the currently unknown location of an impact crater.

REFERENCES AND NOTES
12. All primary data are available as supplementary materials on Science Online.

PROTEIN EVOLUTION

Evolution of protein phosphorylation across 18 fungal species

Romain A. Studer,1,2 Ricardo A. Rodriguez-Mías,2,3 Kelsey M. Haas,2,4 Joanne I. Hsu,2 Cristina Viéitez,3,4 Carme Solé,4 Danielle L. Swanye,2 Lindsay B. Stanford,2 Ivan Latchko,3 René Böttcher,3 Maitreya J. Dunham,2 Eulàlia de Nadal,3 Francesc Posas,3 Pedro Beltrao,1,4 Judit Villén1,4

Living organisms have evolved protein phosphorylation, a rapid and versatile mechanism that drives signaling and regulates protein function. We report the phosphoproteomes of 18 fungal species and a phylogenetic-based approach to study phosphoprotein evolution. We observe rapid divergence, with only a small fraction of phosphosites conserved over hundreds of millions of years. Relative to recently acquired phosphosites, ancient sites are enriched at protein interfaces and are more likely to be functionally important, as we show for sites on H2A1 and eIF4E. We also observe a change in phosphorylation motif frequencies and kinase activities that coincides with the whole-genome duplication event. Our results provide an evolutionary history for phosphosites and suggest that rapid evolution of phosphorylation can contribute strongly to phenotypic diversity.

Divergence in transcriptional regulation and gene expression is often considered the main driver of phenotypic variation across species (I). Comparative gene expression and chromatin immunoprecipitation sequencing studies show that transcriptional regulation has diverged very quickly (2). Similar analysis of protein posttranslational regulation has lagged due to methodological limitations. Now, mass spectrometry (MS) allows comprehensive
identification of posttranslational modification (PTM) events. Studies of phosphosite evolution indicate that phosphorylation constrains protein evolution but that phosphosites can diverge quickly (3–7). However, we lack a phylogenetic understanding of phosphosite evolution. Thus, we examined the regulation of protein function by phosphorylation across 18 different fungal species that diverged from each other tens to hundreds of millions of years ago (Fig. 1A).

Using MS, we identified a total of 73,340 high-confidence (<1% false discovery rate (FDR)) and well-localized phosphosites in these species (ranging from 3092 to 5286 phosphosites per species) (tables S1 and S2) (8). We found that most positions were phosphorylated only in a few species (fig. S1A), partly due to incomplete experimental coverage. As expected, the conservation of the phosphorylation state decreases with the evolutionary distance (fig. S1B).

We estimated the most likely origin for each phosphosite, taking into account the MS-derived phosphosites, sequence-based prediction of phosphorylation potential, and the phylogenetic tree (Fig. 1B) (8). We identified 44,877 origins of phosphorylation in 41,832 aligned positions (Fig. 1C and table S3). Only 890 sites (2.0%) are likely to have been present in the last common ancestor of all 18 species ≤731 million years (My), Y0), and 2194 sites (4.9%) are estimated to be ≤494 My old (Y0+Y1). These ancient sites tend to be conserved in higher eukaryotes, with 39.8% of the Y0 sites and 34.3% of the Y0+Y1 sites also phosphorylated in human or mouse. Phosphosites tend to be more conserved than expected by random sampling of phosphoacceptor residues (fig. S2).

For the 3884 phosphosites identified in *Saccharomyces cerevisiae* (Fig. 1D, left), 69% have an estimated age of ≤18 My, and 90% of sites are ≤182 My old. Age assignment could be biased due to site localization errors or neutral variation whereby the phosphorylation might occur in a neighboring, but not perfectly aligned, residue in an orthologous sequence. To take these into account, we obtained age estimates using a window of ±3 positions around each phosphosite (8). As expected, this analysis shifted the estimated older ages (Fig. 1D; right; fig. S3; and table S3). Using both approaches as bounds, we estimated that 39 to 69% of the *S. cerevisiae* sites are likely to be ≤18 My old, whereas 76 to 90% of sites are likely to be ≤182 My old. This suggests that the majority of observed phosphosites in extant species are evolutionarily novel and change at rates comparable to those observed for transcription factor binding sites (9). Similar results were obtained when restricting the analysis to aligned regions of higher confidence, when mimicking a decrease in coverage, and when restricting the analysis to ancient proteins (figs. S3 and S4) (8). The latter suggests that phosphosites tend to be acquired via mutations in preexisting proteins. We also observed that coverage is not strongly biased by the environmental conditions (fig. S5).

By studying the phylogenetic history of ancient sites with known function in *S. cerevisiae*, we find cases such as T210 in the activation loop of Snf1p kinase, a phosphosite that is critical for viability and is strictly conserved (Fig. 1E). Alternatively, we find cases such as S22 of pyruvate kinase, which, in *S. cerevisiae*, regulates the catalytic activity of this enzyme (J0) but which has been lost more than once during evolution (Fig. 1F). Plausibly, losses like these are compensated for by other mutations and may allow us to study neutral evolution at the level of phosphoregulatory networks. Alternatively, these phosphosites may be important only in specific conditions, resulting in these species occupying environmental niches where the losses are evolutionarily neutral. Out of 61 ancient sites with known function in *S. cerevisiae*, 17 (27.9%) are not phosphoacceptors in at least one species (fig. S6). Out of these cases, four were mutated to negatively charged residues that may mimic phosphorylation and result in constitutive regulation. Evolutionary transitions between phosphosites and acidic residues
have been reported (11) but were not commonly observed in our analysis (8).

Phosphosites with known functions show above-average conservation (6, 12). We therefore hypothesized that phosphosite age may correlate with functional importance. To test this hypothesis, we analyzed the age of S. cerevisiae sites at 2208 protein-protein interfaces, under the assumption that these could regulate protein interactions. We noted that there is a small but significant depletion of young phosphosites at interfaces relative to other exposed residues, supporting the view that sites with a putative function tend to be older (Fig. 2A; test of proportion, P = 9.5 × 10^{-5}). This difference is not observed when normalizing for sequence conservation. Older interface phosphosites have, on average, a stronger predicted destabilizing effect compared with younger sites (Fig. 2A; Mann-Whitney, P = 0.036). This suggests that older interface sites are more likely to be at positions that contribute to the stability of the interface. We further characterized ancient site S28 in S. cerevisiae CDC33 (eIF4E), predicted to be at the interface with TIF4F631 (eIF4G1) (Fig. S7). We purified the CDC33 wild type (WT) and phosphodeficient [S28→A28 (S28A)] and phosphomimetic [S28E] mutants from yeast extracts and quantitatively compared interacting proteins (table S4A) (8). Several observations reveal the functional relevance of this phosphosite. First, both eIF4G paralogs TIF4F631 and TIF4F632 showed a strong increase (four- and twofold, respectively) in binding to CDC33 S28E and the WT versus CDC33 S28A. Second, S28A and S28E mutations moderately compromise eIF4E binding to EAP1, a well-known eIF4E-interacting protein that competes with eIF4G. Third, several proteins involved in stress granule and P-body formation (ASC1, STM1, and TIF34) showed substantially higher affinity for the S28E mutant than the S28A mutant (Fig. 2B). Fourth, the S28E mutant showed increased phosphorylation of neighboring sites at T15 and T22 on CDC33 (fig. S7 and table S4B). These data suggest that CDC33 S28 phosphorylation modulates eIF4E-eIF4G interaction and is a hotspot for translation initiation regulation.

We compared the age of S. cerevisiae phosphosites with functional annotations with those at polymorphic positions. We observed that functionally important phosphosites tended to be older relative to the polymorphic phosphosites (Fig. 2C). Age is predictive of functional phosphosites, even when accounting for sequence conservation (fig. S8), and can therefore be used to guide functional studies. As examples, we studied four ancient phosphosites: H2A1 S121 (S122 counting start-ino), ZUO1 S50, INP53 S975 (Fig. 2D), and CDC10 T216. Compared with the WT controls, INP53 S975A showed an increased resistance to hydroxyurea (HU); ZUO1 S50A showed increased sensitivity to heat; and CDC10 T216 showed no phenotype under osmotic, heat, or HU stress (8). The alanine mutant of H2A1 (S121A), implicated in chromosome stability in S. pombe (13), rendered cells thermosensitive and osmosensitive compared with the WT on both solid and liquid media (Fig. 2C and fig. S9). In nonstressed conditions, the WT and S121A have a very similar expression pattern, with only 32 genes significantly altered in the mutant (FDR < 0.05 and at least two-fold difference) (fig. S9B). Under stress conditions, a larger fraction showed altered gene expression (more than 600 and 800 genes in osmosestress and heat stress, respectively). The expression of repre sentative stress-responsive genes was validated by Northern blot (fig. S9C) and a single cell reporter system (fig. S9D). When stratefying stress-responsive genes by their WT response, genes that were stress-induced in the WT showed a decrease in up-regulation in the mutant, whereas genes that were repressed in the WT showed the opposite behavior (Fig. 2E). These results indicate that S121 on H2A1 is important for maintaining transcriptional control of stress-responsive genes.

The phylogenetic history of phosphosites was used to identify functions and complexes that are enriched in phosphosites of specific age groups (Fig. 3A). To guarantee that the enrichment is due to the age of sites and not the proteins, we normalized the enrichment for the age and average conservation level of the proteins in the corresponding group of genes (8). Several central processes, such as translation and protein folding, as well as some metabolic pathways and some central complexes are enriched in ancient sites (Y0, Fig. 3A). This indicates that the phosphoregulation of these processes and complexes is constrained. Conversely, some processes and complexes are enriched in recently arisen sites that could be...
associated with recent adaptations (Y3 and S1-S4 in Fig. 3A). Different spliceosome components show enrichment for sites of different ages. This potential change in the regulation is consistent with divergence in splicing across these species (14).

We next investigated whether the rapid changes in phosphorylation is also accompanied by changes in the overall usage of phosphorylation motifs and the repertoire of protein kinases (Fig. 3B). We matched each phosphosite to common kinase motif preferences and calculated the relative motif usage across species (8). We observed a trend of lower relative usage of proline-directed motifs and higher relative usage of acidic motifs for species belonging to the post-WGD clade (15) and, in particular, for species in the *Saccharomyces sensu stricto* genus (Fig. 3B, columns one to four).

The differential usage of phosphorylation motifs is not due to a differential presence of the equivalent sequence motifs (figs. S10 and S11), so we speculated it could coincide with changes in the repertoire of protein kinases, their specificities, and/or their activities. We examined the fraction of kinases in each species assigned to different major groups of protein kinases (Fig. 3B, columns 5 to 13, and table S5). We observed a small increase in the fraction of CMGC and STE families in the *Saccharomyces* species and AGC and CAMK kinases in *Saccharomyces* species, Candida glabrata, and Vanderwaterzyma polyspora. However, kinase family membership is a poor predictor of kinase specificity. Thus, we focused on proline-directed (SP) kinases that can be accurately predicted from the post-WGD clade (Fig. 3B, column 14), with the relative presence of SP phosphosites showing a significant correlation with the fraction of SP kinases across species (Fig. 3C; Pearson’s correlation coefficient $r^2 = 0.34, P = 0.011$).

To measure changes in basal kinase activities, we conducted KAYAK (kinase activity assay for kinome profiling) analysis (17). Cell lysates were reacted with peptide libraries matching proline-directed, acidic, and basic motifs. The extent of phosphorylation was used as a proxy for kinase activity toward these sequences (Fig. 3D). We selected nine species spanning our tree and performed two biological replicates ($r^2 = 0.98; $fig. S12A (8). The measured kinase activities show a weak but significant correlation with the relative usage of phosphorylation motifs (Pearson’s $r^2 = 0.09, P = 0.024$; fig. S12B). Relative to the others, *Saccharomyces* species show a significant difference in kinase activities across the three motifs types (analysis of variance, $P = 8.0 \times 10^{-5}$). They tend to have lower activity toward proline and basic motifs and higher activity toward acidic motifs (Fig. 3D).

Together, these results suggest that changes in the repertoire of kinases have altered the overall basal kinase activities, resulting in a differential usage of phosphorylation motifs across species.

On the whole, our results show that evolution of phosphorylation is highly dynamic, and we suggest that changes in PTM regulation are on par with changes in transcriptional regulation in their capacity to quickly generate a diversity of solutions during evolution.

REFERENCES AND NOTES

8. See supplementary materials on Science Online.

*Acknowledgments*

MS data and identifications are accessible via MasseIVE (http://massive.ucsd.edu). RNA sequencing data are deposited in the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo, accession number GSE50258). This work was supported by an Ellison Medical Foundation Award (AG-05-095324-JV), an Arggen grant (to K.M.H.), the Mary Gates scholarship (to J.J.M.H.), and the NSF grant 1143710 (to M.J.D.). Human Frontier Science Program award CDAM0069/2013 (to P.B.) and the European Research Council starting grant ERC-2004-STG-638884 – PoisFunc (to P.B.). M.J.D. is a Rita Allen Foundation Scholar and a Senior Fellow in the Genetic Networks program at the Canadian Institute for Advanced Research. F.P. and E.d.N. are recipients of ICREA (Institució Catalana de Recerca i Estudis Avançats) Acadèmia and are supported by Fundació Botín, Banco Santander, the Spanish government (BFU2016-74297-C3-3-P, BFU2016-73233-P, FEDER), and the Catalan government (2014-SGR-599). We thank the laboratory of S. Fields for providing yeast strains and C. Landry and P. Flicek for critical reading of the manuscript.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6309/229/suppl/DC1

Materials and Methods

Supplementary Text

Figs. S1 to S12

Tables S1 to S5

References (J8–58)

8 January 2016; accepted 15 August 2016

10.1126/science.aaf2144
Editor's Summary

Phosphorylation and fungal evolution

Phosphorylation after transcription modifies the activity of proteins. To understand how phosphorylation sites have evolved, Studer et al. studied a range of fungal species (see the Perspective by Matalon et al.). Only a few sites were apparently present in the common ancestor of all 18 species investigated. Evolutionary age appeared to predict the potential functional importance of specific conserved phosphosites.

Science, this issue p. 229; see also p. 176