tyrosine kinases in signaling, cell-cell communication, and tissue boundary formation (4, 5). We observed a striking negative correlation of genomically encoded tyrosine content with the number of distinct cell types in metazoan species (Spearman’s ρ = −0.89, approximate P = 3.0 × 10^{-9}; Pearson’s r = −0.89, approximate P = 4.0 × 10^{-9}) (Fig. 1A). Thus, metazoans with more cell types have proportionally less potential tyrosine phosphosites. Similarly, we observed that the number of tyrosine kinase domains correlates negatively with genomic tyrosine content (Spearman’s ρ = −0.68, approximate P = 3.7 × 10^{-4}; Pearson’s r = −0.81, approximate P = 1.3 × 10^{-4}) (Fig. 1B). Including dual-specificity mixed-lineage kinases (MLKs) and mitogen-activated protein kinase kinases (MEKs) revealed a similar pattern (fig. S1A).

These observations suggest an evolutionary model in which the acquisition of a tyrosine kinase results in systems-level adaptation to remove deleterious phosphorylation events that cause aberrant cellular behavior and diseases (4). Assuming that a cell begins with a single tyrosine kinase, which is subsequently duplicated, it follows that the kinases may functionally diverge, as a result of relaxation in evolutionary constraints, to phosphorylate new substrates. Emerging kinase specificities could be retained if new substrates confer selection advantage. However, it is unlikely that every new phosphorylation event is beneficial. We hypothesize that optimization of newly emerged signaling networks would follow (6) through the elimination of detrimental phosphorylation events by tyrosine-removing mutations. Even if many new phosphorylation sites are not deletional, an organism with minimized noisy signaling systems is likely to have a fitness advantage. This scenario is repeated with the subsequent duplication of tyrosine kinases leading to more tyrosine residues lost (7).

Despite several recent systematic phosphoproteomic studies (8), many human proteins have no observed phosphosites. Our model suggests that tyrosine loss had occurred predominantly in these proteins in order to minimize tyrosine phosphorylation. To test this hypothesis, we investigated differences in tyrosine loss between these proteins (Non-pTyr) and those that are tyrosine phosphorylated (pTyr). Comparing members of these two groups to their orthologous proteins in S. cerevisiae (7), which lack conventional tyrosine kinases, enabled us to assess the degree of tyrosine loss that may be triggered by the onset of phosphotyrosine signaling in metazoans.
A significantly smaller fraction of amino acids are tyrosines in human proteins than in their yeast orthologs (approximate $P = 3.5 \times 10^{-5}$, paired Wilcoxon signed rank test) (Fig. 1C). However, this phenomenon was statistically more pronounced in non-pTyr proteins than in pTyr proteins (approximate $P = 5.1 \times 10^{-9}$, Mann-Whitney test) (Fig. 1C). A similar trend was observed on the basis of absolute tyrosine residue counts (approximate $P = 2.0 \times 10^{-7}$, Mann-Whitney test) (Fig. 1C). A similar trend was observed on the basis of absolute tyrosine residue counts (approximate $P = 1.3 \times 10^{-7}$, Mann-Whitney test).

Thus, tyrosine loss was strongly favored in human protein evolution, most notably in protein subsets that are not known to be tyrosine-phosphorylated. Genetic drift (9) is unlikely to account for these differences observed in a large number of evolutionarily distant human-yeast protein orthologs. Because tyrosine is an essential and the most expensive amino acid to biosynthesize (10) after tryptophan and phenylalanine, essentiality and biosynthetic cost could be major factors in the observed loss. However, this is unlikely because we observed a strong positive correlation of number of cell types with tryptophan and a weaker negative correlation for phenylalanine (table S1). Instead, we propose that positive selection of tyrosine-removing mutations occurred in the metazoan lineage to reduce adventitious tyrosine phosphorylation, at least in part. This optimization process probably shaped signaling networks crucial for the development of multicellular animals. Additionally, this could provide a mechanism to prevent unspesific phosphorylation events that operate with the evolution of domains, consensus motifs (11), and contextual factors to colocalize kinases with their substrates (11–13). We observed a slightly stronger negative correlation of genomically encoded tyrosine content with the number of inferred phosphotyrosine-binding domains than tyrosine kinase domain count (Spearman’s $\rho = -0.81$, Pearson’s $\rho = -0.88$) (fig. S1A), which is in agreement with the notion that tyrosine phosphorylation exerts parts of its functional effects through creating binding sites for phosphotyrosine domains like Src homology 2 (SH2) and phosphotyrosine binding domain (PTB) (14).

The choanoflagellate Monosiga brevicollis, which is a member of the only known unicellular lineage with canonical tyrosine kinases (15), is an outlier in the cell-type correlation studied above. This observation is consistent with the emerging picture that choanoflagellates represent a distinct evolutionary branch from metazoans in which phosphotyrosine-signaling systems have been used for divergent functions (16, 17). Nevertheless, the Monosiga analysis is still consistent with optimization of phosphotyrosine signaling in this lineage; compared with the metazoans analyzed here, Monosiga has higher numbers of tyrosine kinases.

![Fig. 1. Correlation of expansion of phosphotyrosine-signaling systems with loss of genome-encoded tyrosine residues.](image-url)

(A) The genomically encoded tyrosine content in metazoan organisms and yeast correlate negatively and significantly with organism complexity as measured by distinct cell types (2). Bakers’ yeast (S. cerevisiae) is included as a unicellular eukaryote for comparison. The species analyzed are yeast (S. cerevisiae), worm (C. elegans), sea squirt (C. intestinalis), fly (D. melanogaster), mosquito (A. gambiae), zebrafish (D. rerio), tetraodon pufferfish (T. nigroviridis), Japanese pufferfish (T. rubripes), frog (X. tropicalis), chicken (G. gallus), dog (C. familiaris), cow (B. taurus), mouse (M. musculus), rat (R. norvegicus), chimpanzee (P. troglodytes), and human (H. sapiens). (B) The number of tyrosine kinase domains in metazoans and yeast correlates negatively and significantly with the number of distinct cell types. (C) The fraction of tyrosines in human-yeast ortholog protein pairs. Every point in the scatter plot represents a human-yeast ortholog protein pair where the $(x, y)$ values denote the tyrosine content in human and yeast proteins, respectively. For simplicity, only proteins with an inferred one-to-one orthologous relationship between human and yeast are analyzed (for example, to avoid accelerated sequence divergence due to functional redundancy of paralogs). Orthologous protein pairs lying above the red diagonal ($x = y$) lines have higher tyrosine content in yeast than in human. The left scatter plot is for 437 human proteins conserved in yeast and known to be tyrosine phosphorylated, and the right plot is for 647 human proteins conserved in yeast not known to be tyrosine phosphorylated.
Evolution of a Novel Phenolic Pathway for Pollen Development

Michiyo Matsuno,1* Vincent Compagnon,2* Guillaume A. Schöch,2 Martine Schmitt,3 Delphine Debayle,1 Jean-Etienne Bassard,4 Brigitte Pollet,5 Alain Hehn,5 Dimitri Heintz,1 Pascaline Ullmann,1 Catherine Lapierre,6 François Bernier,1 Jürgen Ehlting,6 Danièle Werck-Reichhart1*

Metabolic plasticity, which largely relies on the creation of new genes, is an essential feature of plant adaptation and speciation and has led to the evolution of large gene families. A typical example is provided by the diversification of the cytochrome P450 enzymes in plants. We describe here a retroposition, neo-functionalization, and duplication sequence that, via selective and local amino acid replacement, led to the evolution of a novel phenolic pathway in Brassicaceae. This pathway involves a cascade of six successive hydroxylations by two partially redundant cytochromes P450, leading to the formation of N2,N2′-dihydroxyferuloyl)-N10-sinapoylsermidine, a major pollen constituent and so far-overlooked player in phenylpropanoid metabolism. This example shows how positive Darwinian selection can favor structured clusters of nonsynonymous substitutions that are needed for the transition of enzymes to new functions.

We recently performed a global coexpression analysis to predict the function of orphan cytochrome P450 genes in Arabidopsis thaliana (5). A candidate emerged with a predicted function in a new branch of phenolic metabolism. CYP98A8 (AT1G74540) and its paralog CYP98A9 (AT1G74550) are two chromosome 1–clustered duplications of an ancestor of CYP98A3, the latter previously shown to meta-hydroxylate the p-cumaric esters of shikimic/quinic acids to form lignin monomers (6, 7). CYP98A8 and CYP98A9 share only about 50% protein identity with CYP98A3, but they do belong to the monophyletic CYP98A clade that includes all confirmed 4-coumaroyl shikimate/quinate meta-hydroxylases; but in protein phylogenies, CYP98A8 and CYP98A9 appear separated from vascular plant sequences by CYP98s from conifers and moss (fig. S1B). This means either that CYP98A8 and CYP98A9 diverged before angiosperm evolution or that they appeared recently and evolved fast, and were thus placed at the base of the clade because of long-branch attraction. In favor of the latter hypothesis, cDNA phylogenies show that CYP98A8 and CYP98A9 form a sister group with CYP98A3 (Fig. 1A and fig. S1D), and no potential orthologs of CYP98A8 exist in the TIGR Plant Transcript Assemblies and PlantGDB databases, except for a single transcript assembly (CYP98A43) from Brassica napus. CYP98A8 and CYP98A9 are intronless, whereas all other known higher plant CYP98As contain two introns at conserved positions, including one that is considered a signature of the expanded CYP71 clade to which the CYP98 family belongs (8). This indicates a gene birth event via mRNA-mediated transposition.

Ratios of nonsynonymous and synonymous substitution rates (dNS/dS or ω) were first es-

1Institut de Biologie Moléculaire des Plantes, CNRS-UFR2357, Université de Strasbourg, 26 Rue Goethe, 75034 Strasbourg Cedex, France.
2Pharma Research, Discovery Technologies, F. Hoffmann-La Roche, CH-4070 Basel, Switzerland.
3Institut Gilbert Laspretis L1C, UMR CNRS/UPEI 7175, Faculté de Pharmacie de Strasbourg, 74 Route du Rhin, 67400 Illkirch, France.
4UMR 266 Chimie Biologique AgroParisTech/INRA, AgroParisTech, 78850 Thévelier-Gignac, France.
5UMR Nancy Université, INRA Agronomie et Environnement, 2 Avenue Forêt de Haye, Bte Postale 172, 54505 Vandoeuvre-lès-Nancy, France.
6Centre for Forest Biology, University of Victoria, Post Office Box 3020 STN CSC, Victoria, British Columbia V8W 3N5, Canada.
7These authors contributed equally to this work.
8To whom correspondence should be addressed. E-mail: danielle.werck@ibmp-ulp.u-strasbg.fr