Sciencexpress Report

Positive Selection of Tyrosine Loss in Metazoan Evolution

Chris Soon Heng Tan,[1,2,3] Adrian Pasculescu,[1] Wendell A. Lim,[4] Tony Pawson,[1,2]* Gary D. Bader,[1,2,3]* Rune Landing[5]*

[1]Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada. [2]Department of Molecular Genetics, University of Toronto, Toronto, Canada. [3]Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Canada. [4]Howard Hughes Medical Institute and Department of Cellular and Molecular Pharmacology, University of California, San Francisco, USA. [5]Cellular and Molecular Logic Team, Section of Cell and Molecular Biology, The Institute of Cancer Research (ICR), SW3 6JB, London, UK.

*To whom correspondence should be addressed. E-mail: pawson@lunenfeld.ca (T.P.); gary.bader@utoronto.ca (G.D.B.); linding@icr.ac.uk (R.L.)

John Nash showed that within a complex system individuals are best off if they make the best decision that they can, taking into account the decisions of the other individuals. Here, we investigate if similar principles influence the evolution of signaling networks in multicellular animals. Specifically, by analyzing a set of metazoan species, we observe a striking negative correlation of genomically encoded tyrosine content with biological complexity (as measured by the number of cell types in each organism). We discuss how this observed tyrosine loss correlates with the expansion of tyrosine kinases in the evolution of the metazoan lineage and how it may relate to the optimization of signaling systems in multi-cellular animals. We propose that this phenomenon illustrates genome-wide adaptive evolution to accommodate beneficial genetic perturbation.

It is a biological paradox that organism complexity shows limited correlation with gene repertoire size (*1*). However, some protein families (*2*) have expanded with organism complexity as measured by number of cell types (*3*), especially those involved in regulation, such as tyrosine kinases in signaling, cell-cell communication and tissue boundary formation (*4, 5*). We observe a striking negative correlation of genomically-encoded tyrosine content with the number of distinct cell types in metazoan species (Spearman's $\rho = -0.89$, approximate P-value = 3.0×10^{-6}; Pearsons $\rho = -0.89$, approximate P-value = 4.0×10^{-6}, Fig. 1A). Thus, metazoans with more cell types have proportionally less potential tyrosine phospho-sites. Similarly, we observed that the number of tyrosine kinase domains correlates negatively with genomic tyrosine content (Spearman's $\rho = -0.68$, approximate P-value = 3.7×10^{-3}; Pearson's $\rho = -0.81$, approximate P-value = 1.3×10^{-4}, Fig. 1B). Including dual-specificity MLK and MEK kinases revealed a similar pattern (fig. S1A).

These observations suggests an evolutionary model where 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 elimination of detrimental phosphorylation events by tyrosine-removing mutations. Even if many new phosphorylation sites are not deleterious, 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 (see SOM).

Despite several recent systematic phospho-proteomic studies (*7*), many human proteins have no observed phospho-tyrosines. Our model suggests tyrosine loss had occurred predominantly in these proteins 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 (see SOM), which lack conventional tyrosine kinases, enabled us to assess the degree of tyrosine loss that may be triggered by the onset of phospho-tyrosine signaling in metazoans.

A significantly smaller fraction of amino acids are tyrosines in human proteins than in their yeast orthologs (approximate $P = 3.5 \times 10^{-4}$, 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 based on absolute tyrosine residue counts (approximate $P = 2.0 \times 10^{-7}$, Mann-Whitney test, fig. S1B), and on a higher confidence subset of pTyr proteins that either have multiple phospho-tyrosines or have sites observed in multiple studies (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 (*8*) is unlikely to account for these differences observed in a large number of evolutionarily distant human-yeast protein orthologs. As tyrosine is an essential and the most expensive amino acid to biosynthesize (*9*) after tryptophan and phenylalanine, essentiality and biosynthetic cost could be major factors in the observed loss. This is unlikely however, because we observed a strong positive correlation of number 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 likely shaped signaling networks crucial for the development of multi-cellular animals. Additionally, this could provide a mechanism to prevent unspecific phosphorylation events, that operates with evolution of domains and contextual factors to co-localize kinases with their substrates (*10*, *11*). Tyrosine phosphorylation typically exerts its functional effects through allosteric regulation, or by creating binding sites for phospho-binding domains like SH2 and PTB (*12*). In agreement, we observed a slightly stronger negative correlation of genomically-encoded tyrosine content with the number of inferred phospho-tyrosine binding domains than tyrosine kinase domain count (Spearman's $\rho = -0.81$, Pearson's $\rho = -0.88$, see fig. S1A).

We note that the choanoflagellate Monosiga brevicollis, which is a member of the only known unicellular lineage with canonical tyrosine kinases (*13*), is an outlier in the cell type correlation studied above (data not shown). This observation is consistent with the emerging picture that choanoflagellates represent a distinct evolutionary branch from metazoans in which phospho-tyrosine signaling systems have been used for divergent functions (*14*, *15*). Nevertheless, the Monosiga analysis is still consistent with optimization of phospho-tyrosine signaling in this lineage – compared to metazoans analyzed here, Monosiga has higher numbers of tyrosine kinases and lower genomically-encoded tyrosine content (data not shown).

Other factors, such as tyrosine sulfation, could have contributed to the observed tyrosine loss, which raises the question whether other post-translational modifications and regulatory mechanisms are under similar evolutionary selection. We observed strong negative correlation of number of cell types with amino acids that can be methylated or glycosylated (table S1). The numbers of genomically-encoded threonine showed strong negative correlations with serine/threonine kinase and cell type numbers, although these trends were not observed with serine (fig. S2), suggesting possible coarse-grained functional differences between serine-and threonine-phosphorylation in metazoans.

Our findings suggest that the implementation of tyrosine kinase signaling, as a biological innovation that likely assisted the development of multi-cellular organisms, required system-level adaptive mutations. Analogous to the arguments by John Nash in his dissertation (*16*), this phenomenon highlights a general principle of adaptive evolution pertaining to the introduction of new components into a complex system, and parallels evolution of some human societies where the local populations have to adjust and adapt to the influx of immigrants contributing to the societies' economic development. This principle may serve as an important framework when considering the evolution and fidelity of complex biological systems. Finally, this work raises the possibility that complex regulatory diseases, such as cancer, might result from systems-wide adaptive changes in human genomes and signaling systems.

References and Notes

1. E. Szathmry, F. Jordn, C. Pl, *Science* **292**, 1315 (2001).
2. C. Vogel, C. Chothia, *PLoS Comput Biol* **2**, e48 (2006).
3. J. T. Bonner, *Integ. Biol.* **1**, 28 (1998).
4. T. Hunter, *Curr Opin Cell Biol* (2009).
5. Z. Songyang, L. C. Cantley, *Trends Biochem Sci* **20**, 470 (1995).
6. A. Zarrinpar, S.-H. Park, W. A. Lim, *Nature* **426**, 676 (2003).
7. C. Jørgensen, R. Linding, *Brief Funct Genomic Proteomic* **7**, 17 (2008).
8. S. Wright, *Genetics* **16**, 97 (1931).
9. D. W. Raiford, et al., *J Mol Evol* **67**, 621 (2008).
10. R. Linding, et al., *Cell* **129**, 1415 (2007).
11. M. L. Miller, et al., *Sci Signal* **1**, ra2 (2008).
12. B. T. Seet, I. Dikic, M.-M. Zhou, T. Pawson, *Nat. Rev. Mol. Cell Biol.* **7**, 473 (2006).
13. N. King, et al., *Nature* **451**, 783 (2008).

14. D. Pincus, I. Letunic, P. Bork, W. A. Lim, *Proc Natl Acad Sci U S A* **105**, 9680 (2008).

15. G. Manning, S. L. Young, W. T. Miller, Y. Zhai, *Proc Natl Acad Sci U S A* **105**, 9674 (2008).

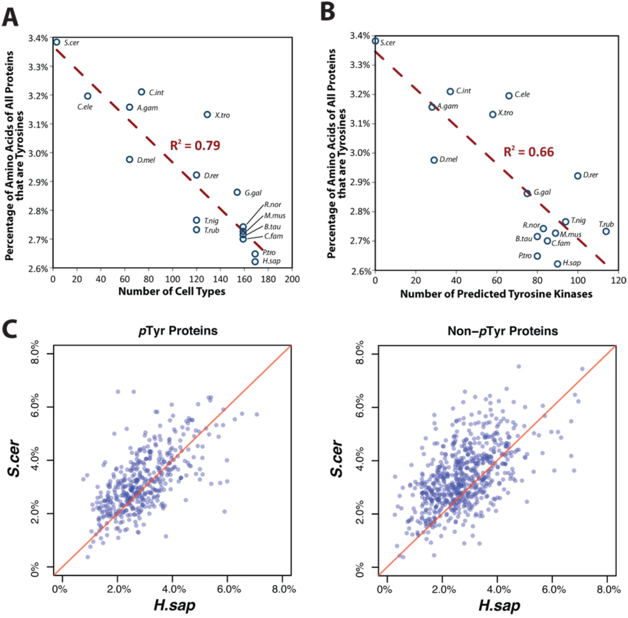
16. J. Nash, Non-cooperative games, Ph.D. thesis, Princeton University (1950).

17. We thank Claus Jørgensen, Jiangzhi Zhang, Karen Colwill, Jing Jin and Kresten Lindorff-Larsen for suggestions and fruitful discussions. This project was in part supported by Genome Canada through the Ontario Genomics Institute and the Canadian Institutes of Health Research (MOP-84324). C.S.H.T. conceived the project. C.S.H.T., W.A.L., G.D.B., T.P. and R.L. designed the experiments. C.S.H.T., R.L. and A.P. performed the experiments. C.S.H.T., G.D.B., W.A.L., T.P. and R.L. wrote the paper. R.L. managed the project.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1174301
Materials and Methods
Figs. S1 and S2
Table S1
References and Notes

31 March 2009; accepted 25 June 2009
Published online 9 July 2009; 10.1126/science.1174301
Include this information when citing this paper.

Fig. 1: Correlation of expansion of phospho-tyrosine signaling systems with loss of genome encoded tyrosine residues. 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 composition in yeast than 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.


A
Percentage of Amino Acids of All Proteins that are Tyrosines
3.4%
3.3%
3.2%
3.1%
3.0%
2.9%
2.8%
2.7%
2.6%
S.cer
C.ele
C.int
A.gam
X.tro
D.mel
R² = 0.79
D.rer
G.gal
R.nor
M.mus
B.tau
C.fam
T.nig
T.rub
P.tro
H.sap
0
20
40
60
80
100
120
140
160
180
200
Number of Cell Types
B
Percentage of Amino Acids of All Proteins that are Tyrosines
3.4%
3.3%
3.2%
3.1%
3.0%
2.9%
2.8%
2.7%
2.6%
S.cer
C.int
C.ele
A.gam
X.tro
D.mel
R² = 0.66
D.rer
G.gal
R.nor
T.nig
T.rub
B.tau
M.mus
C.fam
P.tro
H.sap
0
20
40
60
80
100
120
Number of Predicted Tyrosine Kinases
C
pTyr Proteins
S.cer
8.0%
6.0%
4.0%
2.0%
0%
0%
2.0%
4.0%
6.0%
8.0%
H.sap
Non–pTyr Proteins
S.cer
8.0%
6.0%
4.0%
2.0%
0%
0%
2.0%
4.0%
6.0%
8.0%
H.sap