

The evolution of PTM sites in proteins: phosphorylation and N-glycosylation

Summary:

Vikas Rao Pejaver

Lab meeting – 04/19/2011

PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites

Florian Gnäd, Shubin Ren, Juergen Cox, Jesper V Olsen, Boris Macek, Mario Oroshi and Matthias Mann

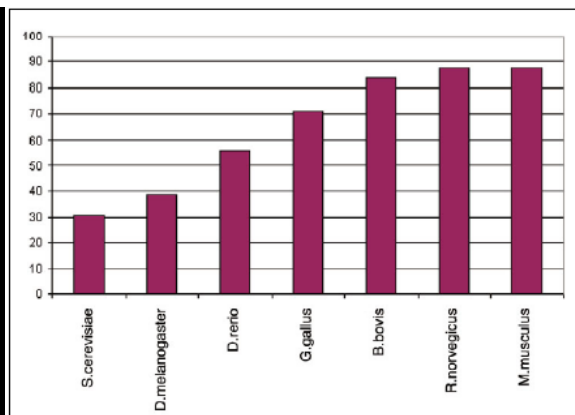


Figure 7

Percentage sequence identity of phosphoproteins with orthologs.

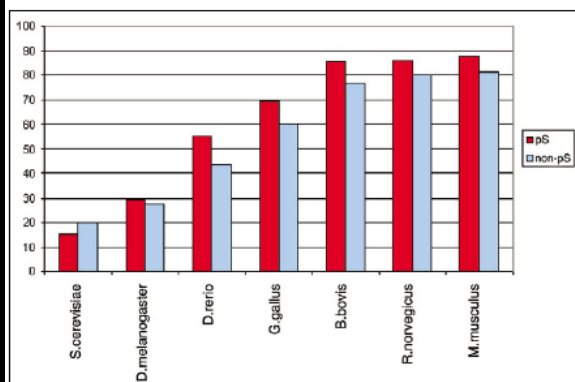


Figure 8

Conservation of phosphoserines (red) compared to non-phosphoserines (blue) in phosphoproteins. Phosphoserines are significantly more conserved except in yeast.

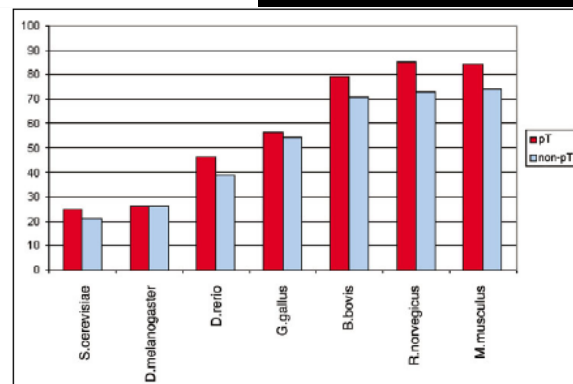


Figure 9

Conservation of phosphothreonines (red) compared to non-phosphothreonines (blue). Phosphothreonines are significantly more conserved within mammals.

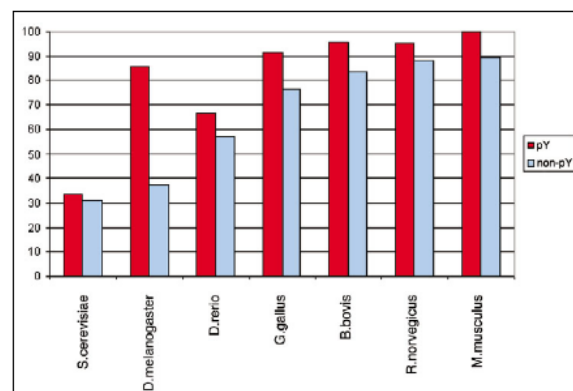


Figure 10

Conservation of phosphotyrosines (red) compared to non-phosphotyrosines (blue). Tyrosine is very highly conserved in mammals in both forms. In more distantly related species the numbers are small and differences are not statistically significant.

What we knew as of 2007...

- ◎ Phosphorylation sites generally more conserved than non-phosphosites (anecdotal evidence and Gnad et al. 2007)
- ◎ Perhaps due to important functional roles that these sites play (more orthologs)
- ◎ Phosphotyrosine conservation is higher in general (Gnad et al. 2007)
- ◎ Gnad et al. also showed -5 and +5 residues from sites conserved
- ◎ Phosphosites prefer disordered regions (Iakoucheva et al. 2004, Dunker et al. 2002, Jimenez et al. 2007)

Comparative phosphoproteomics reveals evolutionary and functional conservation of phosphorylation across eukaryotes

Jos Boekhorst*, Bas van Breukelen[†], Albert JR Heck[†] and Berend Snel^{*‡}

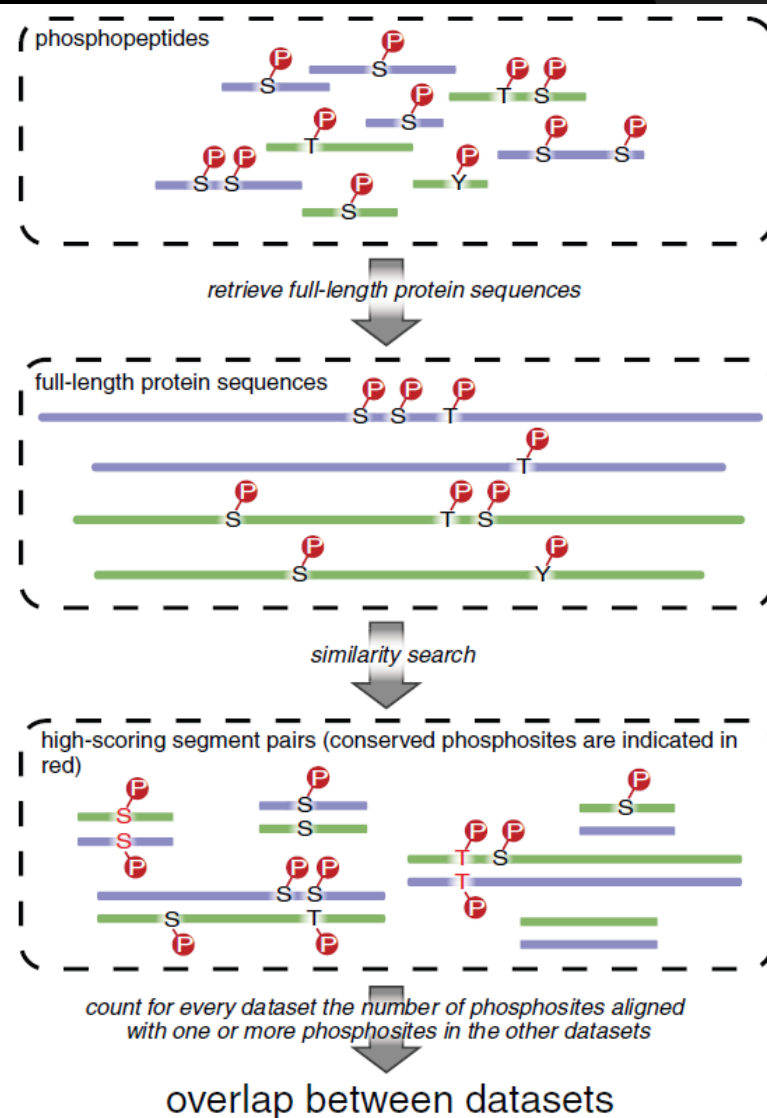
Published: 1 October 2008

Genome Biology 2008, 9:R144 (doi:10.1186/gb-2008-9-10-r144)

Compared phosphoproteomics datasets of six eukaryotes

Phosphoproteomics datasets

Species	Reference	Proteins (n) ^a	Sites (n) ^a
Human	[34]	1,419	3,296
Mouse	[23]	1,605	3,142
Fly	[14]	991	2,080
Yeast	[35]	481	850
Plant	[22]	470	724
Zebrafish	[11]	668	759



Comparative phosphoproteomics reveals evolutionary and functional conservation of phosphorylation across eukaryotesJos Boekhorst*, Bas van Breukelen[†], Albert JR Heck[†] and Berend Snel^{*‡}

Published: 1 October 2008

Genome Biology 2008, 9:R144 (doi:10.1186/gb-2008-9-10-r144)

- ◎ Overlap in datasets ranging from ~700 sites (human and mouse) to one site (fish and yeast)

Number of sites found in three or more different species						
	Three different species ^a		Four different species		Five different species	
	Observed	Expected ^b	Observed	Expected	Observed	Expected
Plant	9	1.55	2	0.02	2	0.00
Fly	33	7.55	13	0.18	2	0.00
Human	103	59.45	17	1.31	3	0.01
Mouse	106	63.59	23	1.80	4	0.02
Yeast	2	0.27	0	0.00	1	0.00
Fish	72	40.15	12	1.86	2	0.01

- ◎ Relative overlap between species reflects actual phylogeny
- ◎ Conservation signal observed despite heterogeneity of the original studies

What do we know now ...

- ◎ Phosphorylation sites prefer disordered regions
- ◎ Generally conserved in a majority of proteins
- ◎ The number of phosphosites shared much greater than expected - some shared by plants and animals (over a billion years)

Non-functional phosphorylations?

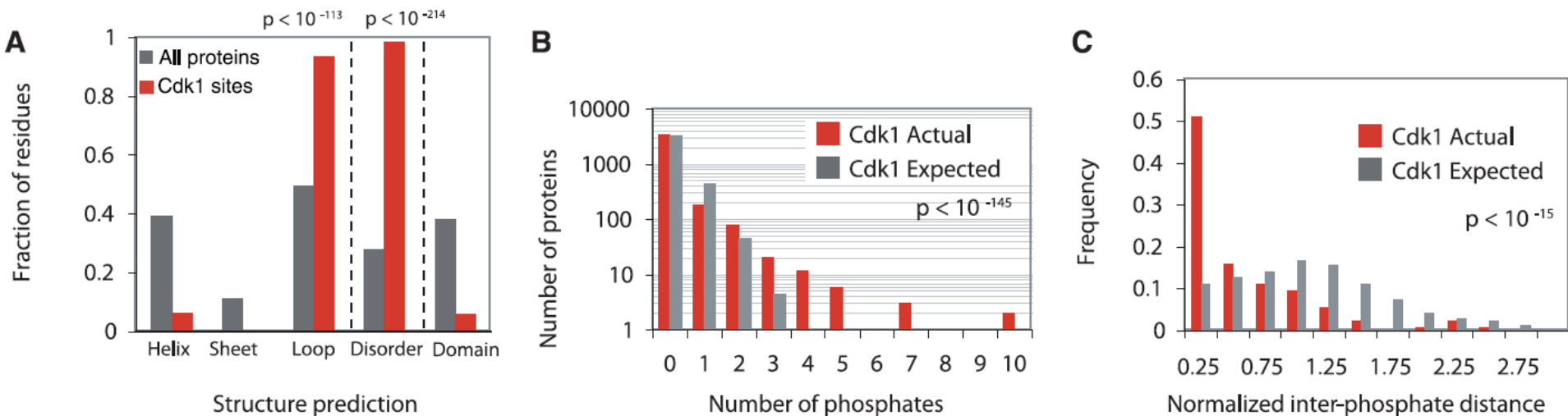
Gustav E. Lienhard

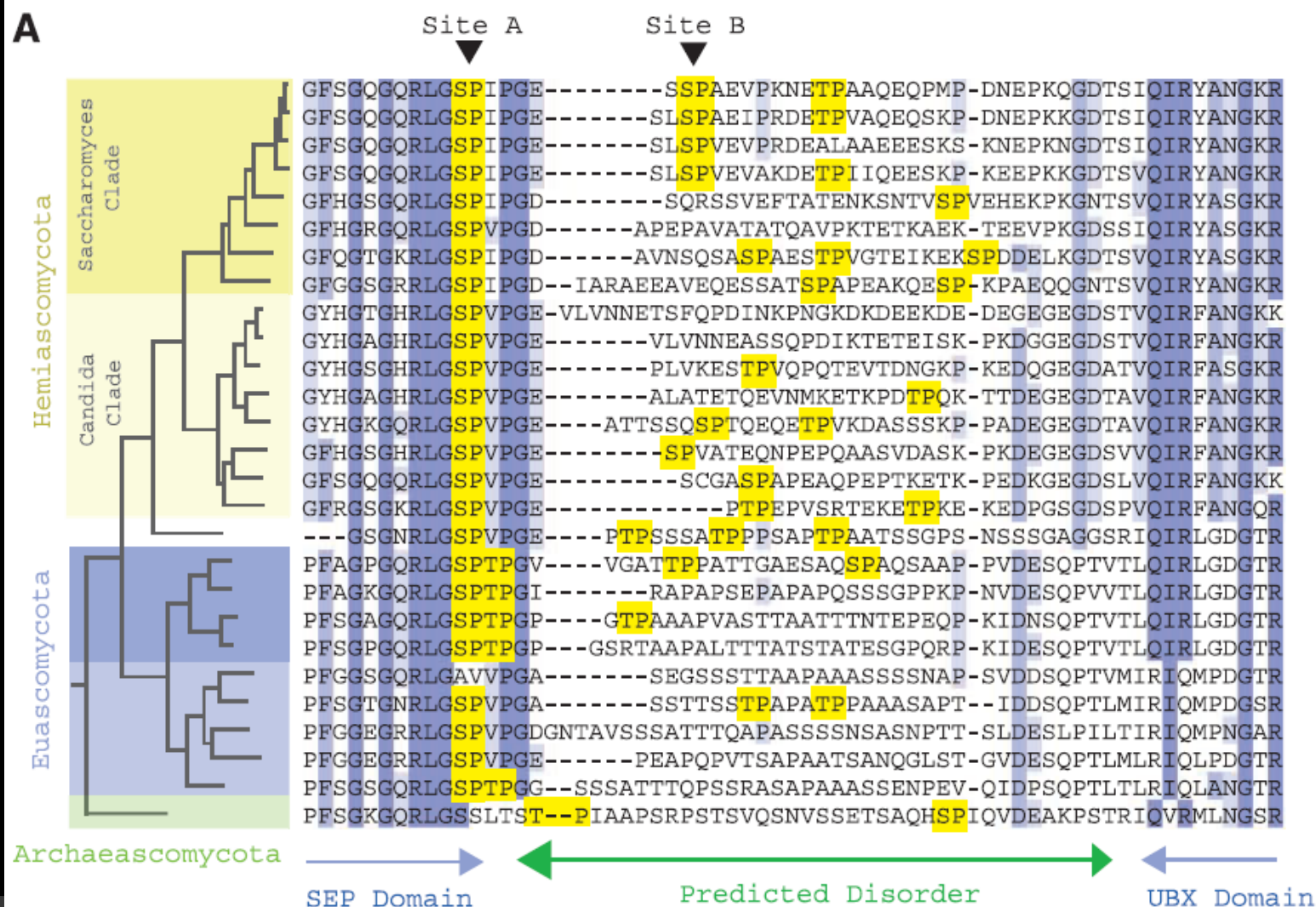
- ⊙ “This proposition is one that has often been discussed informally among scientists in the area of signaling but not yet, to my knowledge, in print.”
- ⊙ Promiscuous kinases – may not have evolved ‘perfect’ substrate-specificity
Neutral consequence - phosphorylation without function may persist if it is not disadvantageous
- ⊙ Bioenergetic argument – non-functional phosphorylation wastes ATP
But ~1% of the cellular ATP production
- ⊙ Stoichiometric argument – non-functional phosphorylations more probable for low stoichiometry cases
Negligible effect and thus, not disadvantageous
- ⊙ Gnad et al. paper – significant proportion of sites in human and mouse orthologs not conserved

What do we know now ...

- ◎ Phosphorylation sites prefer disordered regions
- ◎ Generally conserved in a majority of proteins
- ◎ The number of phosphosites shared much greater than expected - some shared by plants and animals (over a billion years)
- ◎ But a substantial fraction unconserved – the idea of non-functional phosphosites and the neutral evolution of these sites

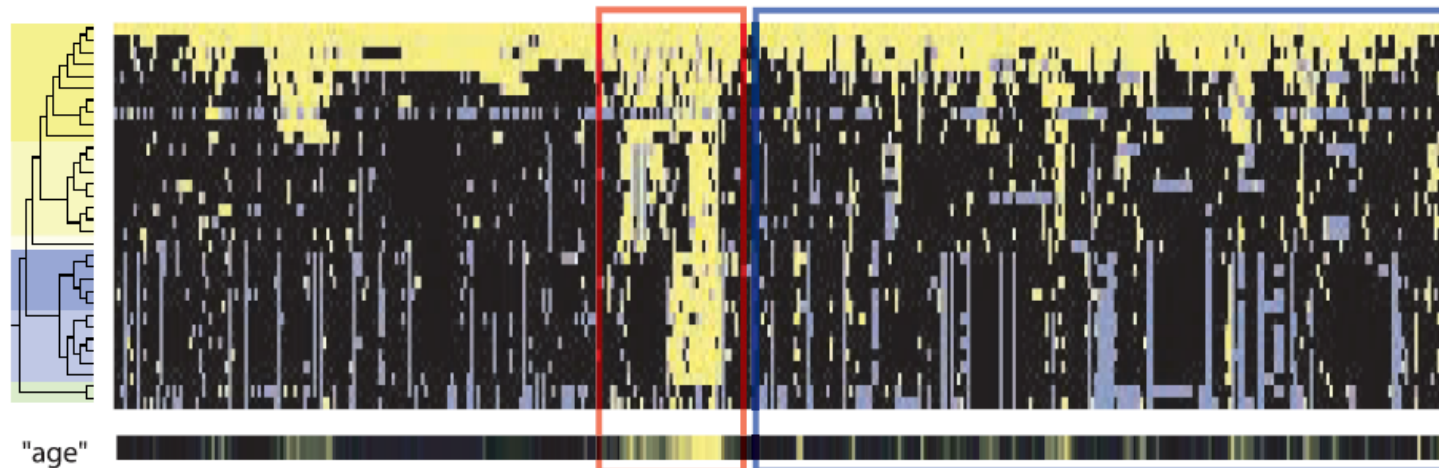
- Identified the positions of 547 phosphorylation sites on 308 Cdk1 substrates in *S. cerevisiae*
- Position of most phosphosites not conserved
- Instead clusters of sites shift position in disordered regions
- Gain and loss - implications for the rapid evolution of kinase-signaling circuits



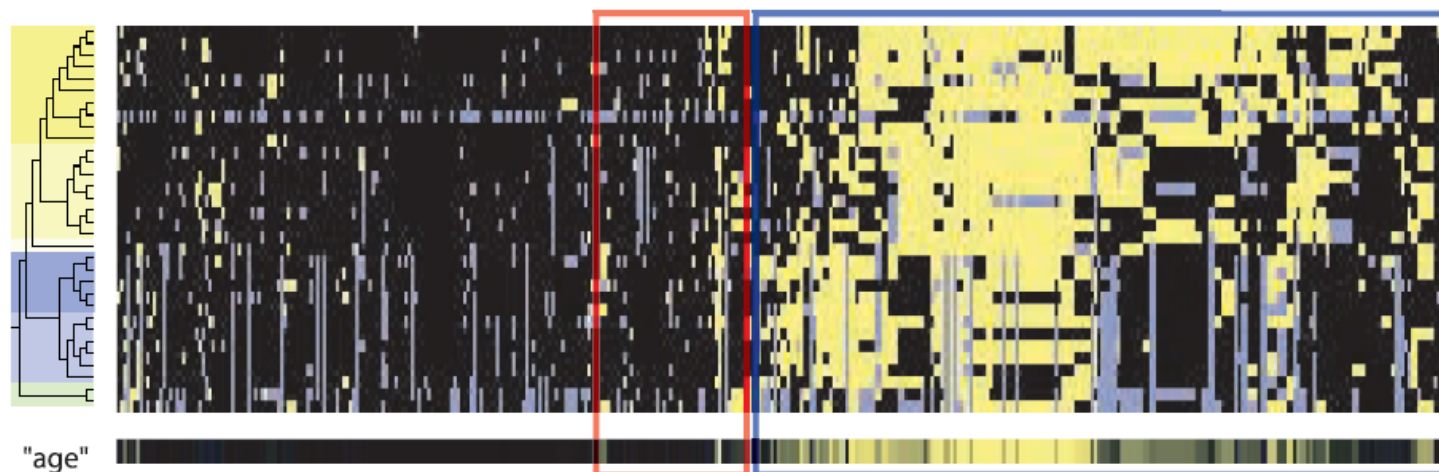


B

conservation of precise position of phosphorylation site



conservation of enrichment of phosphorylation sites within ORF



What do we know now ...

- ◎ Phosphorylation sites prefer disordered regions
- ◎ Generally conserved in a majority of proteins
- ◎ The number of phosphosites shared much greater than expected - some shared by plants and animals (over a billion years)
- ◎ But a substantial fraction unconserved – the idea of non-functional phosphosites and the neutral evolution of these sites
- ◎ Phosphosites tend to cluster together but positions on the protein need not be conserved – shift happens in disordered regions

Weak functional constraints on phosphoproteomes

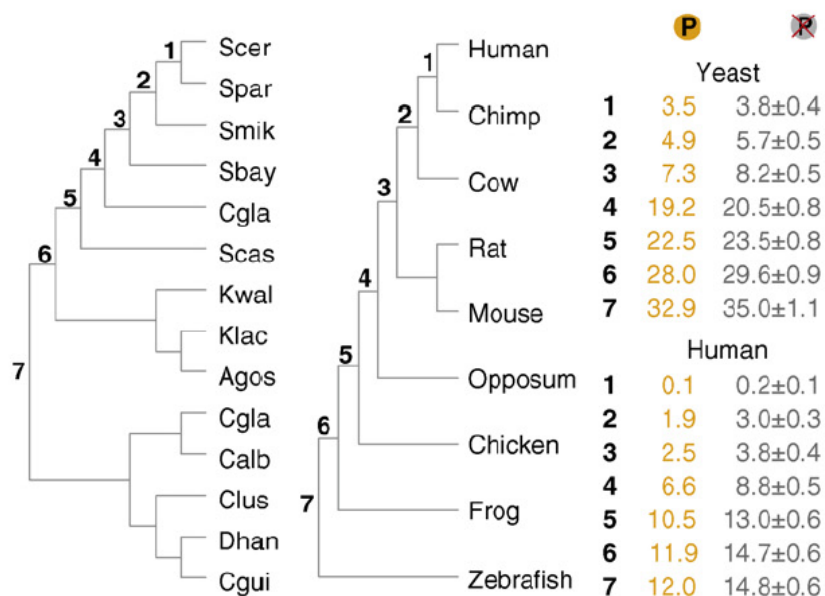
Christian R. Landry^{*}, Emmanuel D. Levy^{*} and Stephen W. Michnick

- Looked at :
 - 5010 sites in 2099 orthologous groups in the yeast lineage
 - 7137 sites in 2347 orthologous groups in the vertebrate lineage
- Reconstructed ancestral sequences by maximum likelihood for each lineage
- Inferred age of phosphosite to get a global picture of the entire phosphoproteome

Weak functional constraints on phosphoproteomes

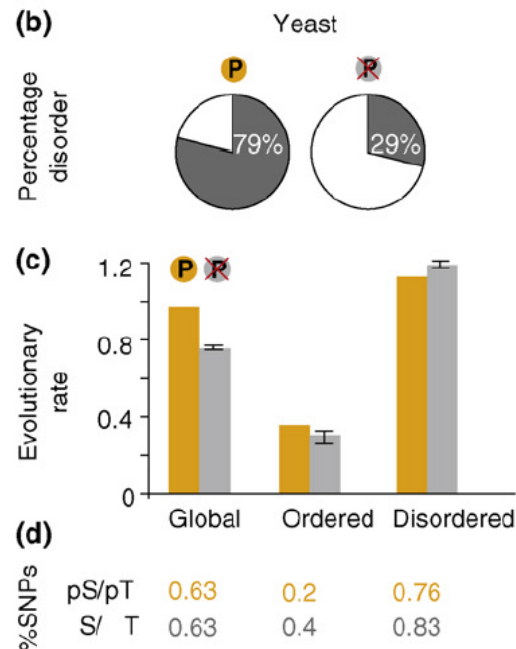
Christian R. Landry*, Emmanuel D. Levy* and Stephen W. Michnick

(a) Percentage residues that appeared at each node



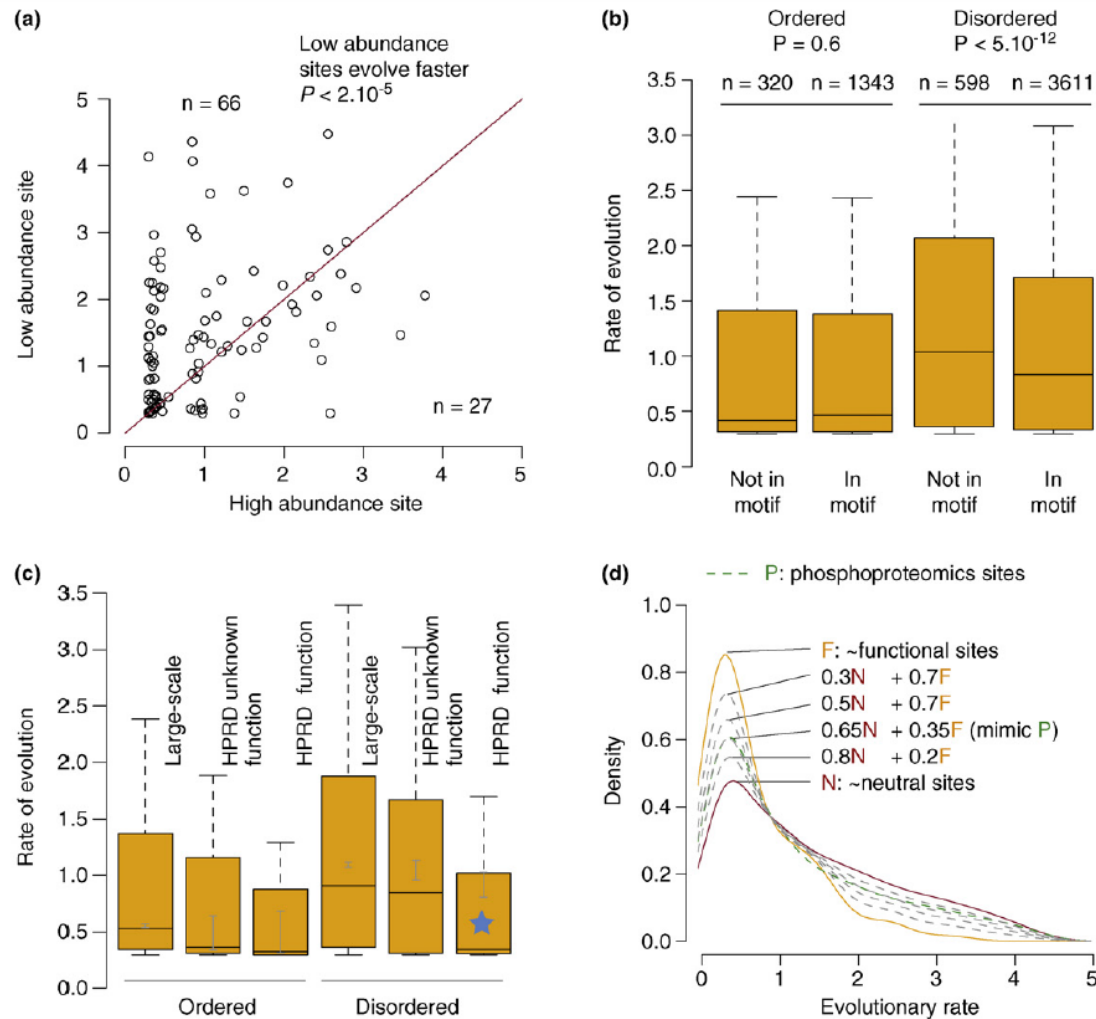
Key: **P** S and T known to be phosphorylated
R S and T not known to be phosphorylated; sampled in same proportion, numbers and proteins

(b)



Weak functional constraints on phosphoproteomes

Christian R. Landry*, Emmanuel D. Levy* and Stephen W. Michnick



The plot thickens ...

- ◎ Phosphorylation sites prefer disordered regions
- ◎ Generally conserved in a majority of proteins
- ◎ The number of phosphosites shared much greater than expected - some shared by plants and animals (over a billion years)
- ◎ But a substantial fraction unconserved – the idea of non-functional phosphosites and the neutral evolution of these sites
- ◎ Phosphosites tend to cluster together but positions on the protein need not be conserved – shift happens in disordered regions
- ◎ Phosphosites on average more conserved when enriched in disordered regions – otherwise weakly conserved (up to 65% can be non-functional)???

Evolution of Characterized Phosphorylation Sites in Budding Yeast

Alex N. Nguyen Ba^{1,2} and Alan M. Moses^{*†,1,2}

- Considered only *S. cerevisiae* and three other species from the same genus
- Sites are significantly constrained when compared with flanking regions (irrespective of disorder)
- Total number of sites within a protein is conserved and cases of compensation observed

Phosphorylated and Nonphosphorylated Serine and Threonine Residues Evolve at Different Rates in Mammals

Sean Chun-Chang Chen,¹⁻⁴ Feng-Chi Chen,⁵ and Wen-Hsiung Li^{*3,6}

- Argue that Landry et al. assumed deep conservation by reconstructing the vertebrate ancestor – so use just mouse, human, dog and opossum (outgroup)
- Phosphorylated Ser and Thr more conserved than non-phosphorylated ones
- Tyr does not follow this pattern – also the human lineage has gained more phosphorylated Thr residues and lost fewer phosphorylated Tyr residues

Evolution of Protein Phosphorylation for Distinct Functional Modules in Vertebrate Genomes

Zhen Wang,^{†,1,2} Guohui Ding,^{†,1,3} Ludwig Geistlinger,⁴ Hong Li,^{1,2} Lei Liu,^{1,3} Rong Zeng^{*,1}
Yoshio Tateno,^{*,5} and Yixue Li^{*,1,3}

- Grouped human phosphosites into functional modules based on KEGG
- Phosphosites more conserved in vertebrate-specific functional modules (cellular signalling and stimuli response) than in basic functional modules (metabolic and genetic processes)

What do we know now ...

- ◎ Phosphorylation sites prefer disordered regions
- ◎ Generally conserved in a majority of proteins
- ◎ The number of phosphosites shared much greater than expected - some shared by plants and animals (over a billion years)
- ◎ But a substantial fraction unconserved – the idea of non-functional phosphosites and the neutral evolution of these sites
- ◎ Phosphosites tend to cluster together but positions on the protein need not be conserved – shift happens in disordered regions
- ◎ Phosphosites on average more conserved when enriched in disordered regions – otherwise weakly conserved (up to 65% can be non-functional)???
- ◎ Phosphosites in disordered regions evolve more rapidly than those in ordered regions?
- ◎ Maybe strongly linked to vertebrate evolution

Rampant Purifying Selection Conserves Positions with Post-Translational Modifications in Human Proteins

MBE

Vanessa E. Gray¹ and Sudhir Kumar^{1,2}

Letter (Revised)

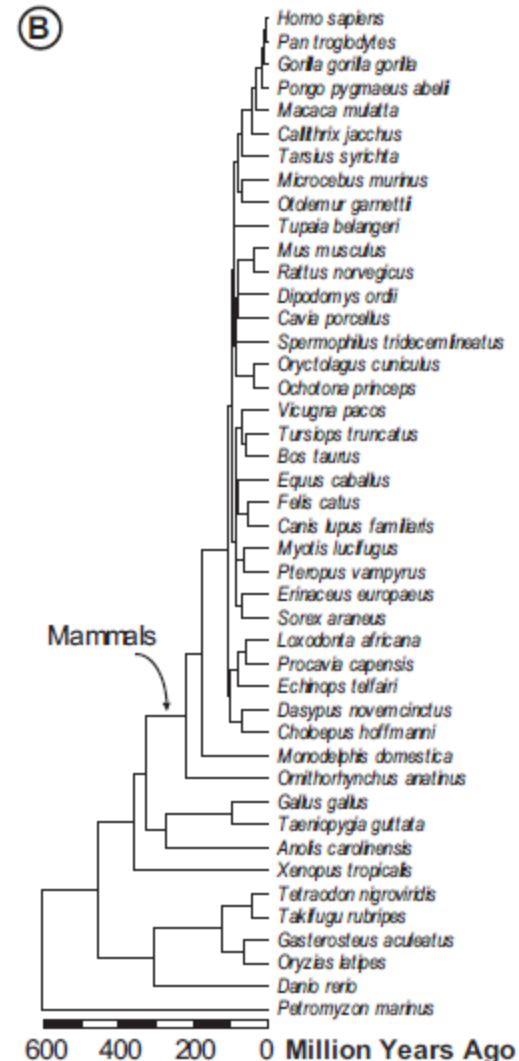
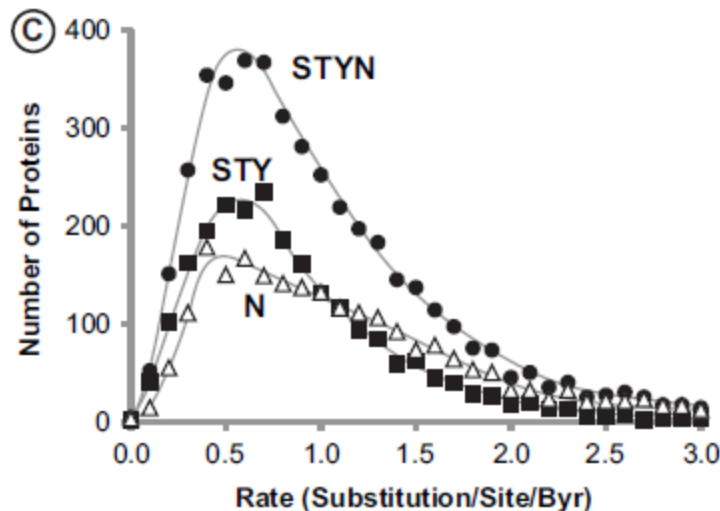
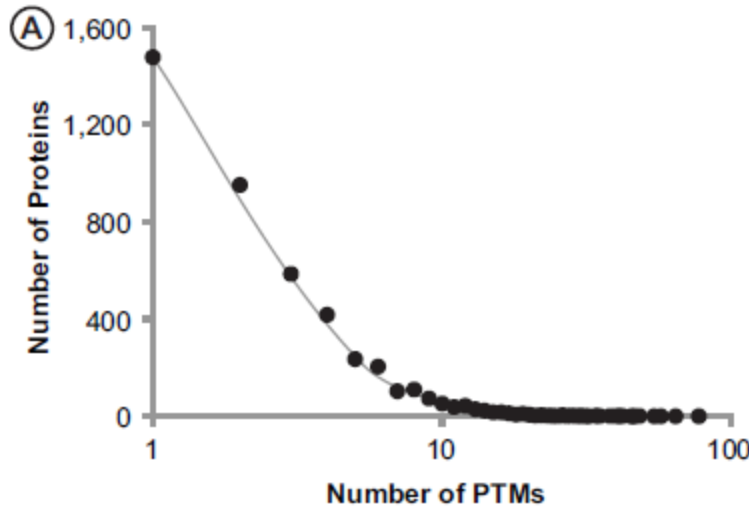
- ⊙ Argue that proteome-wide comparisons neglect evolutionary pressures of individual proteins
- ⊙ Compared 16,836 PTM (phosphorylation and N-glycosylation) positions in 4,484 human proteins
- ⊙ Used 44-species alignment for each protein and mapped each site to the species tree
- ⊙ Calculated absolute evolutionary rates as the ratio of the amount of change to the total time elapsed over all tree branches
- ⊙ 70% PTMS of phosphorylated and N-linked glycosylated proteins show purifying selection
- ⊙ 42% more severe at PTM sites than non-PTM residues
- ⊙ Phosphorylated proteins more conserved than N-glycosylated proteins

Rampant Purifying Selection Conserves Positions with Post-Translational Modifications in Human Proteins

MBE

Vanessa E. Gray¹ and Sudhir Kumar^{1,2}

Letter (Revised)

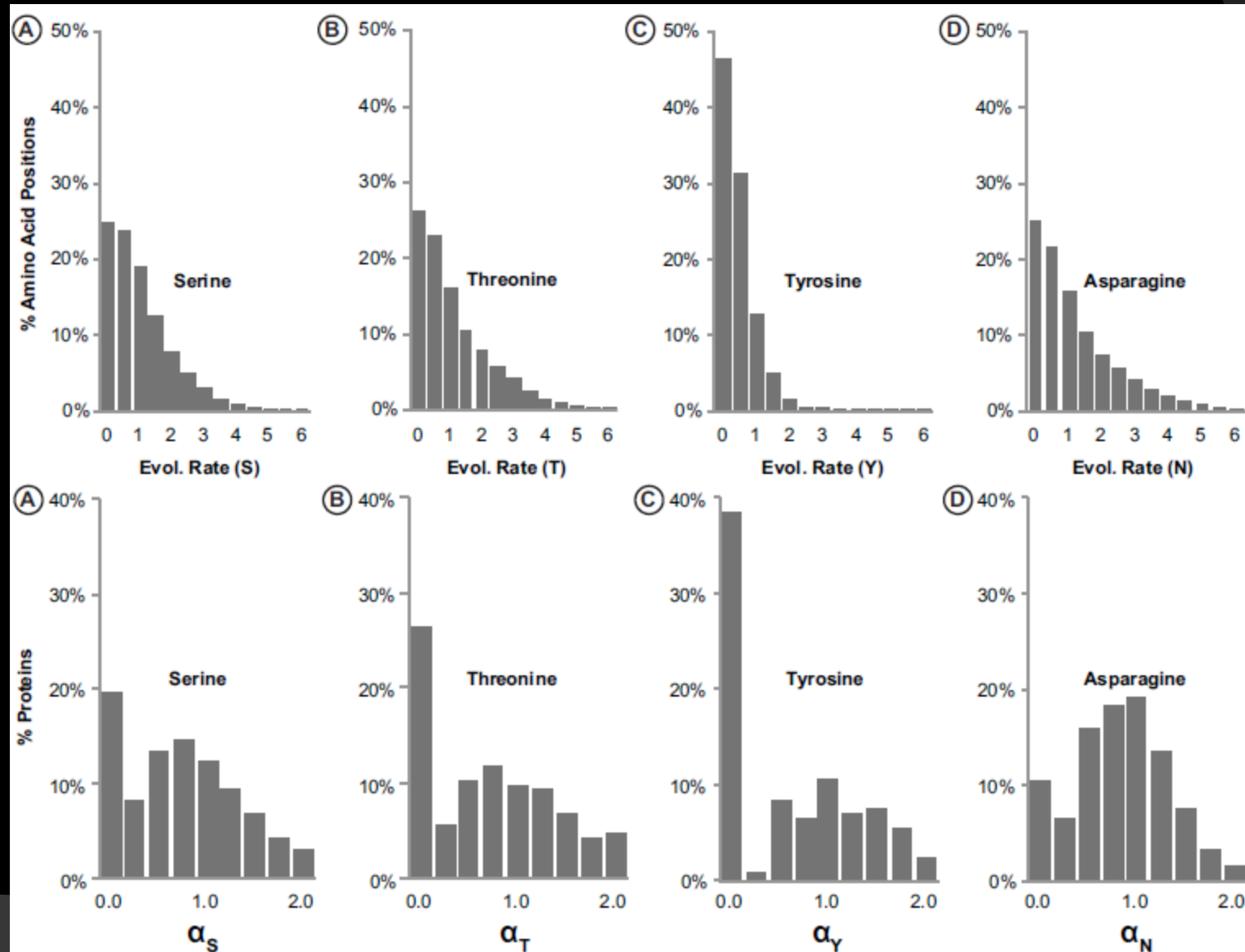


Rampant Purifying Selection Conserves Positions with Post-Translational Modifications in Human Proteins

MBE

Vanessa E. Gray¹ and Sudhir Kumar^{1,2}

Letter (Revised)



What do we know now ...

- ◎ Phosphorylation sites prefer disordered regions
- ◎ Generally conserved in a majority of proteins
- ◎ The number of phosphosites shared much greater than expected - some shared by plants and animals (over a billion years)
- ◎ But a substantial fraction unconserved – the idea of non-functional phosphosites and the neutral evolution of these sites
- ◎ Phosphosites tend to cluster together but positions on the protein need not be conserved – shift happens in disordered regions
- ◎ Phosphosites on average more conserved when enriched in disordered regions – otherwise weakly conserved (up to 65% can be non-functional)???
- ◎ Phosphosites in disordered regions evolve more rapidly than those in ordered regions
- ◎ Maybe strongly linked to vertebrate evolution
- ◎ The idea of weak constraints on phosphosites completely debunked?

Genome-wide evolutionary conservation of N-glycosylation sites

Chungoo Park and Jianzhi Zhang^{*}

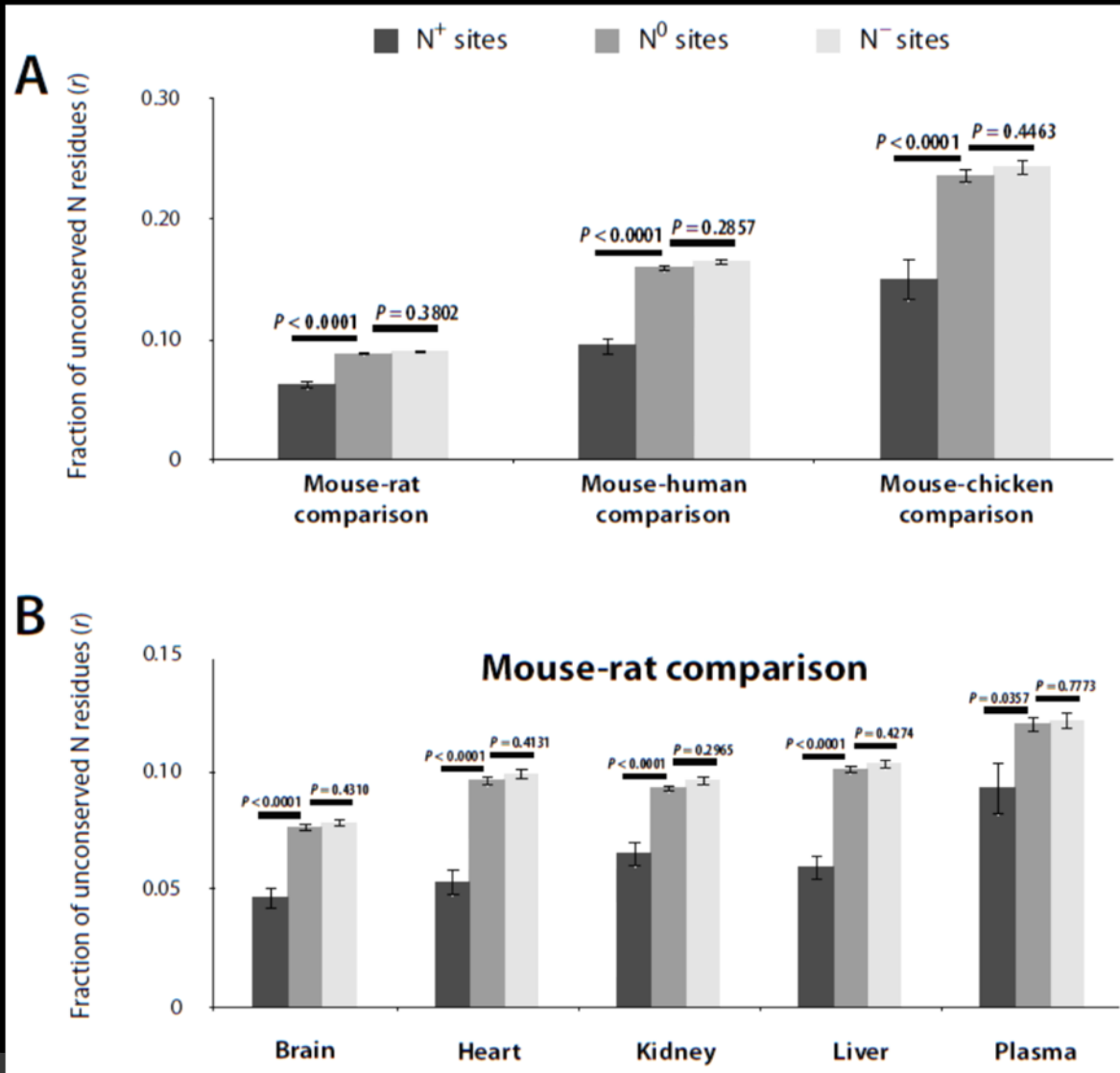
MBE Advance Access published February 25, 2011

- ◎ Looked at mouse N-glycosylation – 6367 sites in 2352 proteins
- ◎ Also compared to 36,000 phosphosites from 6296 mouse proteins
- ◎ Glycosylated Asn evolves more slowly than non-glycosylated Asn
- ◎ Not biased by conservation of the region itself
- ◎ No redundancy among multiple glycosylated sites within a protein

Genome-wide evolutionary conservation of N-glycosylation sites

Chungoo Park and Jianzhi Zhang*

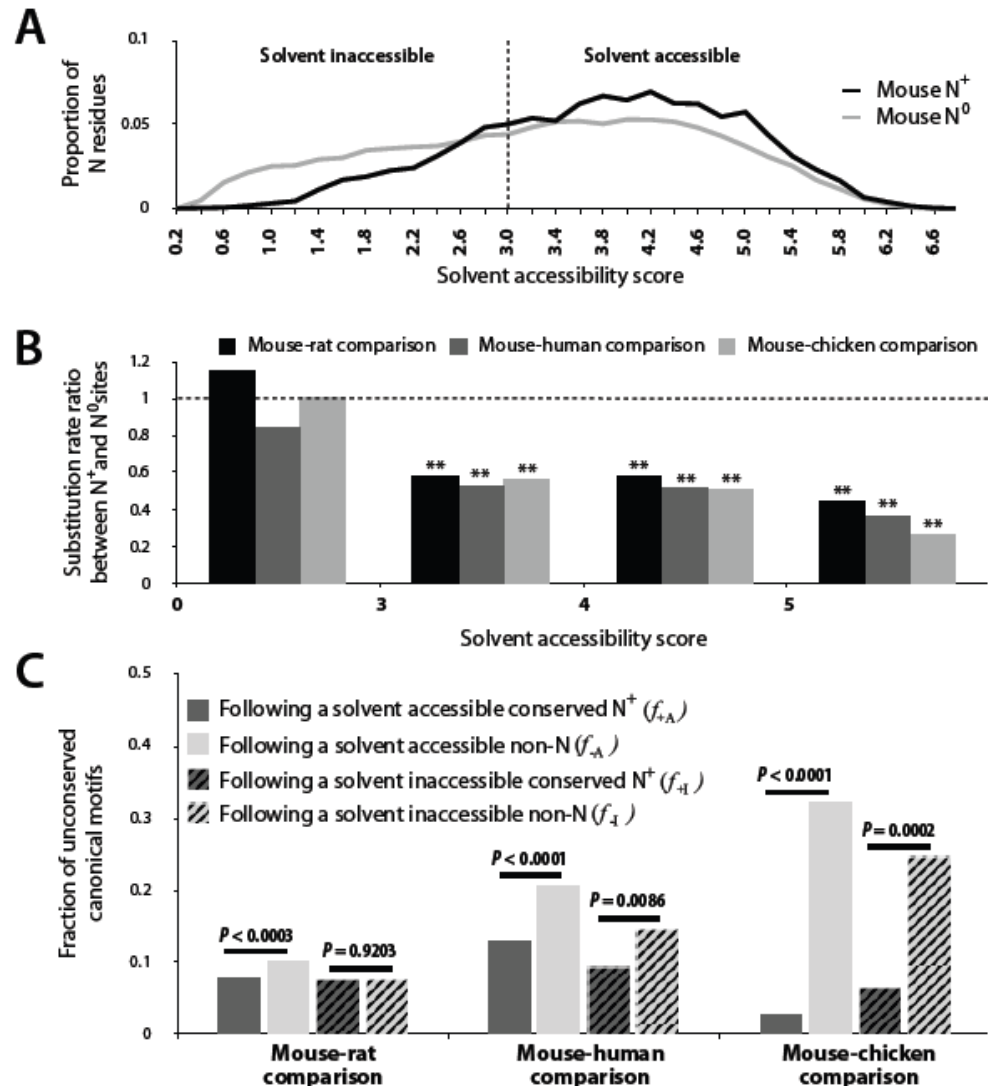
MBE Advance Access published February 25, 2011



Genome-wide evolutionary conservation of N-glycosylation sites

Chungoo Park and Jianzhi Zhang*

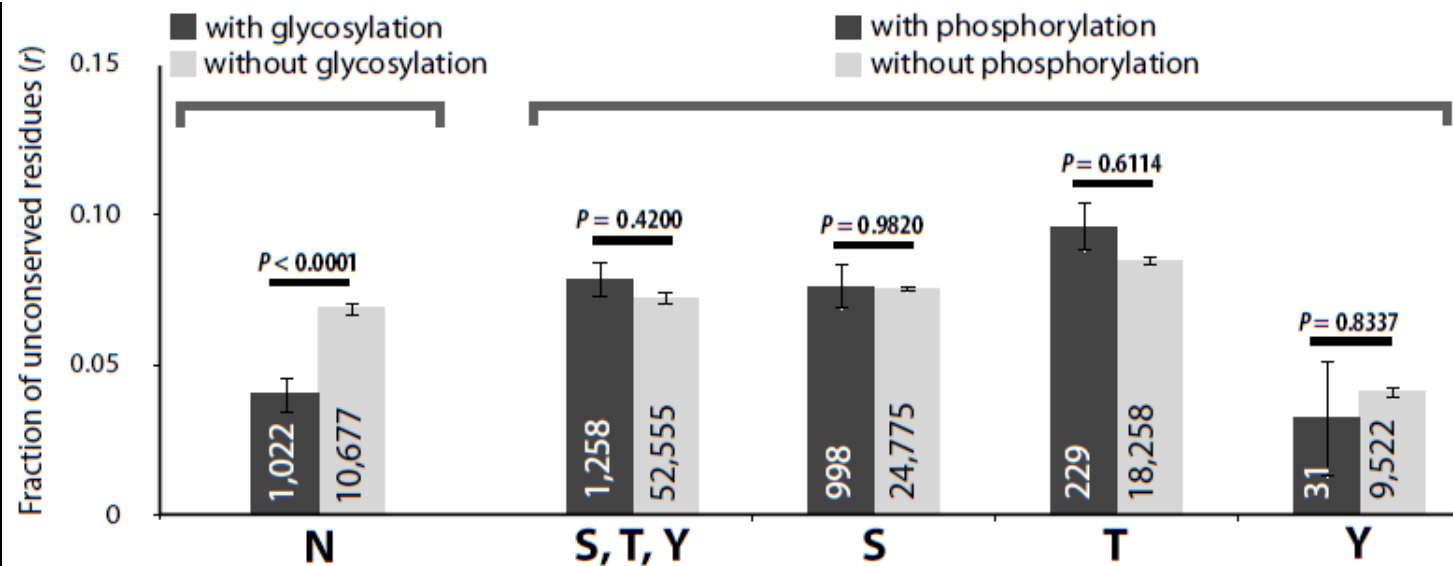
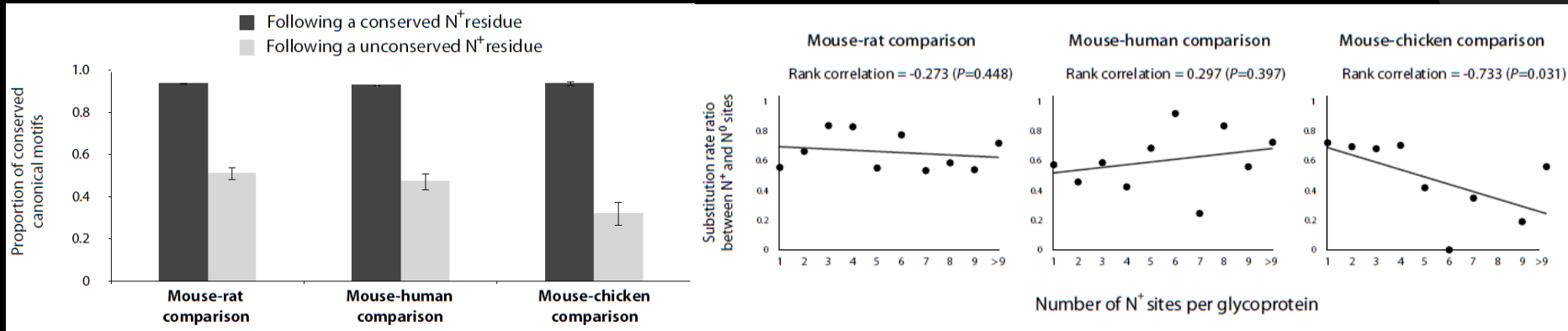
MBE Advance Access published February 25, 2011



Genome-wide evolutionary conservation of N-glycosylation sites

Chungoo Park and Jianzhi Zhang*

MBE Advance Access published February 25, 2011



What do we know today...

- ◎ Phosphorylation sites prefer disordered regions
- ◎ Generally conserved in a majority of proteins
- ◎ The number of phosphosites shared much greater than expected - some shared by plants and animals (over a billion years)
- ◎ But a substantial fraction unconserved – the idea of non-functional phosphosites and the neutral evolution of these sites
- ◎ Phosphosites tend to cluster together but positions on the protein need not be conserved – shift happens in disordered regions
- ◎ Phosphosites on average more conserved when enriched in disordered regions – otherwise weakly conserved (up to 65% can be non-functional)???
- ◎ Phosphosites in disordered regions evolve more rapidly than those in ordered regions
- ◎ Maybe strongly linked to vertebrate evolution
- ◎ The idea of weak constraints on phosphosites completely debunked?
- ◎ N-glycosylation site conservation is not that ambiguous
- ◎ Strong functional constraint on solvent accessible N-glycosylation sites
- ◎ N-glycosylation more conserved than phosphorylation?

Thank you for listening!
My thoughts on all of this?

