# Genes encoding intracellular signaling proteins in animals originated primarily in Opsithokonta and Vertebrata

Floriane Picolo[1], Jérémie Bardin[2], Michel Laurin[2], Benoît Piégu[1], and Philippe Monget[*1]

[1] PCR, UMR85, INRAE, CRNS, IFCE, Université de Tours, F-37380 Nouzilly, France

[2] CR2P "Centre de Recherches sur la Paléo-biodiversité et les Paléo-environnements", UMR 7207, CNRS/ MNHN, Muséum National d'Histoire Naturelle, Sorbonne Université, Paris

Corresponding author: philippe.monge@inrae.fr

## Abstract

In this work, we evaluate the time of appearance of genes involved in animal signaling pathways with the aim of studying the times of appearance of genes relative to their partners. Signaling pathways are very little described outside humans. We therefore used the human signaling pathways described in kegg (47) as a framework to infer the relationships between genes. Furthermore, we searched for the orthologs of these genes in 315 animal species plus yeast, grouped into 25 clades, in order to determine the time of appearance of each gene.

For this study, 47 human intracellular signaling pathways are used, and the animal life tree to the most common ancestor of vertebrates and yeast, with a species panel of 316 animal species and a tree comprising 25 nested clades to determine the timing of gene onset. Our hypothesis is that there is a link between the time of appearance of a gene and its position in a signaling pathway. We first found that 2 key clades stand out for all genes: Opisthokonta and Vertebrata. In a second step, we observe that for interactions 2 to 2 involved in a signaling pathway, only 16% of genes arose simultaneously, compared to 40% of interacting interactants on a partner arrived before their partner and 43% arrived after their partner. In addition, for 25 pathways, we have a negative correlation between the time of birth of genes and their position in the pathway, that is, the earlier the gene originates in the tree of life, the later it interacts in the signaling pathway.

## Introduction

The co-evolution of genes encoding interacting molecules is a subject of intense study (Andreani & Guerois, 2014; Fraser et al., 2004; Lynch & Hagner, 2015; Rand et al., 2004) because of the intriguing question of the modes of mutation and selection that act on two molecules simultaneously. In particular, the co-evolution of the binding motif has been well investigated (Lewis et al., 2010). These studies of co-evolution focused for example on the fitness (Bloom et al., 2004; Williams et al., 2001), on the conservation of the interaction (Kachroo et al., 2015; Lovell & Robertson, 2010; Wuchty et al., 2003), or on the evolution of the residues at the interface of the molecules (Echave & Wilke, 2016; Jack et al., 2016; Mintseris & Weng, 2005). While these studies on the coevolution of binding partners often require the integration of different disciplines (chemistry, evolution, biology), the establishment of the interaction from a phylogenetic point of view is less studied. Little is known for example about the origin and evolution of the different partners prior to their first interaction. Does the emergence of one partner favor the emergence of the second partner?

In the case of interacting molecules, the appearance of genes coding for molecules included in a complex is more intricate (Kauffman, 1993). For two molecules that will eventually interact, the appearance of one may be dependent on the appearance and conservation of the other. This may be the case, for example, when the presence of the first molecule is not advantageous if its partner has not yet appeared (a sort of exaptation).

The existence of interacting proteins without partners ("orphan partners") has been frequently described (Howard et al., 2001), even though it is sometimes difficult to assess whether an interacting protein is a true orphan or its ligand is just unknown (Benoit et al., 2006). The relative order of appearance of genes encoding protein partners is thus an open question. Furthermore, several types of interactions can be observed in living organisms, with different numbers of interacting partners (Albelda & Buck, 1990; Koshland, 1958; Maslov & Sneppen, 2002; Sullivan & Holyoak, 2008), varying affinities (Kent et al., 1980), or different duration for the interactions (Nooren & Thornton, 2003), making the problem more complex.

Concerning the pairs of ligand and its receptor, Thornton (2001) has shown that the first steroid receptor of the family, present in lamprey and supposed to be present in the last common ancestor of vertebrates, was an œstrogen receptor, and that several duplications led to other steroid receptors, specialized in other functions with other ligands. However, more recent investigations suggested that the ancestral ligand for the ancestral steroid receptor was a molecule with a structure distinct from modern œstrogen, an aromatized steroid with a side-chain, called paraestrol (Markov et al., 2017). Concerning membrane receptors, in a previous work, we found that 41% of the receptors and their respective first ligands appeared on the same branch of the evolutionary tree, representing 2.5-fold more than expected by chance, thus suggesting an evolutionary dynamic of interdependence and conservation between these

partners (Grandchamp & Monget, 2018). In contrast, 21% of the receptors appeared after their ligand, i.e., three-fold less often than expected by chance, and 38% of the receptors appeared before their first ligand, as much as expected by chance. These results suggest that a selective pressure is exerted on ligands and receptors once they appear, that would remove molecules whose partner does not appear quickly.

In the present work, we investigated the node of appearance of genes encoding proteins involved in signaling pathways, downstream of membrane receptors in the animal tree of life (Figure 1). We also studied the relationship between the position of the protein in the signaling pathway relatively to his partners, and the node of appearance in the tree of life. In other words, does the order of appearance of genes in the tree of life reflect that of their coded protein in the signaling pathway (upstream or downstream of the pathway), with a direct or reverse order (Figure 2)?

# Results

## The nodes of origin of genes encoding proteins of signaling pathways

First, we looked at the distribution of birth times for each of our genes encoding proteins involved in an animal signaling pathway (2,298 unique genes involved in 47 human intracellular signaling pathways available on the KEGG database V104.0, see *M&M*). Two parts of the tree of life are overrepresented in gene origins (Figure 3): Yeast and Porifera at the base of the tree, which can be more broadly encompassed in *Opisthokonta* (13000 MYA), and the *Tunicata, Petromyzontidae, Mixini , Chondrichthyes and Actinopterygii*, which correspond to the first diversification of Vertebrates (758 MYA) (Kumar et al., 2022).

Moreover, we looked at each pair of proteins interacting with each other to study whether both appeared at the same time/node in the tree of life, or whether one arrived before the other, and if so, the length of the delay. Of the 3,000 interactions studied, 16.3% of interactions are “simultaneous”, i.e., both partners in an interaction appeared at the same clade, and 83.7% are asynchronous, of which 43.2% are “forward”, i.e., partner 2 (downstream in the pathway) arrived first, 40.5% are “backward”, i.e., partner 1 (upstream in the pathway) arrived first (Figure 4). Some pathways have more backward relationships than others, such as the MAPK and IL-17 pathways in comparison with the "Ovarian steroidogenesis" pathway, in proportion to their respective numbers of interactions. It is the opposite for other pathways such as the PPAR and cAMP pathways.

Secondly, simulation procedures allowed us to compare the relative quantities of interaction of each type (forward, backward or simultaneous) to distributions obtained by randomly mixing the relationships between the elements of the pathways. For 4 pathways, Hedgehog, RIG-I-like receptor, C-type lectin receptor and Thyroid hormone (Suppl. Data 1), there are significantly more forward interactions (>56.20%) than if these interactions were taken at random ($p < 0.05$). For 5 other pathways (Sphingolipid, AMPK, Notch, Ovarian Steroidogenesis and Thyroid hormone), backward interactions were found more often (> 50.3%) than randomly ($p < 0.05$), and in particular for the Ovarian Steroidogenesis pathway for which in 91.5% of the tests we found more backward interactions than random ones. Finally, concerning the Hippo pathway, in 84.2% of the tests, we found significantly more interactions appearing simultaneously than by taking interactions at random, and this is the case for 24 other pathways (>53.6%; $p < 0.05$). For 13 pathways, we did not find a significant difference between the chance of establishing one of the three types of interaction (backward, forward or simultaneous) and establishing it at random.

We then looked at the delta between the node of birth of the first interactant and that of the second interactant for the 47 pathways studied (Figure 5). For 12 pathways (at the top of Figure 6), the delta median is greater than 0 (between 0.5 and 3), for 6 pathways the median is less than 0 (between -2 and -1), and for 29 pathways the median is 0. Moreover, certain pathways have specific profiles. Firstly, the

PPAR pathway exhibits a very restricted boxplot with very minimal variance (min = -1, Q1 = -0.50, median = 0, Q3 = 0.5 and max = 1). And some pathways exhibit median particularly eccentric to 0 such as the IL-17 pathway (min = -12, Q1 = -8, median = -2, Q3 = 0.5 and max = 12), and the "ovarian steroidogenesis" pathway (min = -6, Q1 = 0.5, median = 3, Q3 = 5.5 and max = 14) (Q1 and Q3 being 1 and 3rd quantiles of a dataset).

For each pathway, we also studied the delta depending on the node of appearance in the tree of life of each partner. For the Hippo pathway, for example, the distribution curve resembles a normal distribution, the interactions having appeared in the same clade (delta = 0) (13.6%) (to the Porifera clade, clade 2 (45.5% of total interaction with delta = 0, (Figure 6 deciphering the line Hippo of the figure 5)). For the Ovarian steroidogenesis and cAMP pathways, the deltas are greater than 0 (median = 3 and 1, respectively), the last elements of the pathway have arrived before terminal elements (Suppl Data 2). The other pathways are represented in Suppl. data 2.

## Relationship between the node of birth of a partner involved in each pathway, and the upstream or downstream position in the pathway

For 25 of the 47 pathways studied here, there is a negative correlation (p-value range from 3.73E-104 to 0.02) between the node of birth of a gene and the position of the corresponding protein in the pathway, which means that upstream proteins in a signaling pathway tend to appear late in the tree of life, for instance in vertebrates to eutherians (Hypothesis 2 of Figure 2 B, Table 1 and Figure 8). For three of them, this correlation is particularly strong: Adipocytokine (r = -0.58 and p = 2.44E-24), Fc epsilon RI (r = -0.53 and p = 2.02E-16), and VEGF (r = -0.53 and p = 3.15E-8). For 10 pathways, the correlation between the node of appearance and the position in the pathway is positive (p-value range from 2.79E-280 to 0.03), which means that the more upstream the protein is involved in the pathway, the older the corresponding gene is (node of appearance in Opisthokonta and Metazoa for example). All the correlations are presented in Table 1.

The moment of appearance of each member of a pathway on the KEGG pathways shows interesting patterns (Figure 8 and Suppl. Data 4). In the Hippo pathway (Figure 8 A), there are a lot of ancestral members, distributed throughout the pathway, upstream and downstream (hypothesis 3, Figure 2 C). In the Hippo and the Wnt pathways (Figure 8 A), a lot of partners appear between clades 1 and 2 (Opisthokonta or Metazoa): Mob, TEAD (clade 1), Mer, SAV1, Mst1/2 (clade 2) and YAP/TAZ (clade 3). And the downstream members of these pathways (after the nucleus) arose in vertebrates (clade 12). In the PPAR pathway (Figure 8 B), there are 3 central elements, represented by PPARα (clade 12), δ (clade 12) and γ (clade 14), dimerized with RXR (α (clade 13), δ (clade 15) and γ (clade 17)). The PPAR pathway corresponds to hypothesis 2 (Figure 2 B), upstream proteins in the signaling pathway tending to appear late in the tree of life. We also observe examples of our hypothesis 3 (Figure 2 C) with the Notch pathway for example (Figure 8 C), there is not necessarily an order that emerges, and we do not

find any correlation between the positions in the pathway and the clade of birth of corresponding gene. We also have examples of our hypothesis 3 (Figure 2 C) with the Notch pathway for example (Figure 8 C), there is not necessarily an order that emerges, and we do not find any correlation between the positions in pathways and times of gene appearance of proteins in pathways and times of gene appearance ($r = 0.02$, pvalue = 0.77). We also observe that no pathway corresponds to hypothesis 1 (Figure 2 A), upstream proteins in a signaling pathway never appearing late in the tree of life (Suppl.data 4).

# Discussion

In our previous work (Grandchamp & Monget, 2018), we have shown that the genes encoding the pairs of ligand/membrane receptor mainly appeared at the root of metazoa, vertebrates and teleosts. In the present work, we show that the genes encoding proteins of signaling pathways mainly appeared in Yeast and Porifera clade in the one hand, and in the node of Vertebrates on the other hand. Moreover, as expected, among all the pathways, at least 218 proteins (from 1 to 33 by pathway) appeared in yeast, showing, as expected, that numerous genes encoding proteins of signaling pathways appeared long before membrane receptors and their ligands. Our previous study on membrane receptors and their ligands was done on a tree with only 10 nodes, and it could be refined by using the 25 nodes as in the present paper.

In the signaling pathways studied, there are therefore numerous proteins that interact with their partner(s) in humans, appearance of both being desynchronized in the tree of life. This raises the question of the functioning of these pathways without the full set of his members. For example, p53, which appeared in Metazoa, is inhibited by several factors whose genes appeared later, such as Mdm2 and Mdm4 (Olfactora). Without these inhibitors, individuals would die *in utero,* as shown in the mouse (Jones et al., 1995; Montes de Oca Luna et al., 1995). This suggests there was another p53 inhibitor before Olfactora appeared. In addition, human p53 targets also emerged later in evolution, such as IGFBP3 (Vertebrates) and pro-apoptotic factor bax (Olfactora). It is these and other factors that explain the tumor suppressor role of p53 in humans (Lehmann-Che et al., 2007) but also in *Drosophila* (Zhou, 2019). We can therefore hypothesize a progressive refinement of the mechanisms of action and inhibition of p53 during evolution.

Another example is the interaction between PIN1 and IRF3 in the RIG-I-like receptor pathway, which have a delta of 13, PIN1 gene having appeared at clade 1 (Yeast) and IRF3 at clade 14 (Chondrichthyes). Pin1 also activates p53 which appeared on the branch leading to clade 2 (node of Metazoa) (Berger et al., 2005). Another example of desynchronization concerns the Sphingolipid pathway, in which CTSD interacts only with BID, and BID interacts with CTSD and BAX in humans. CTSD appeared in Metazoa, BAX in Olfatora, and BID gene in Prototheria. Moreover, although yeast genome does not contain genes encoding Bcl-2 proteins, the heterologous expression of mammalian Bax in yeast induces a suppressible lethal phenotype that is associated with characteristics of metazoan apoptosis, strongly suggesting that its targets are already present in yeast (Khoury & Greenwood, 2008; Zha et al., 1996).

We found that for 33 (4+5+24) pathways, there are significantly more forward, backward or simulated interactions that have been established than if these interactions were taken randomly. This suggests that for 14 pathways, the directions of these interactions were established by chance. In other words, for these interactions, they seem to have been established by chance, and probably because it worked in the cell, these interactions were conserved during evolution.

In a previous study, we described the cases where the genes coding for membrane receptors appeared before their ligand (Grandchamp & Monget, 2020). We had studied more precisely the 30 cases for which the 3D structures were known in the PDB database, to formulate hypotheses on the plausible scenarios of evolution of the amino acids involved in the binding pocket of the receptor, until the ligand appeared. In the present work, a similar study would be very interesting for the partners which appeared before the proteins with which they interact. Such a study could be performed for ras/sos complexes for example, grb2/sos, akt/gsk3 or akt/mTor, for which the 3D structure is available.

Some signaling pathways are relatively old (i.e., contain more members that appeared in yeast or metazoans than in more recent clades), such as the MAPK pathway, other pathway being more recent, such as the pathways involved in immunity. Of note, genes involved in immune pathways evolve very quickly compared to the rest of the genome (Cooper & Alder, 2006).

Almost all signaling pathways have target factors (e.g., factors entering the nucleus) that appeared very early during evolution (dark blue factors to the right in the colored KEGG diagrams, Figure 8 and Suppl data 3). However, for certain pathways, the factors that appeared earliest (yeast or metazoa) are close to the targets, and those further upstream in the pathway appeared later (PPAR, cAMP), whereas for other pathways (IL-17, T cell receptor), certain factors close to the targets appeared later than the factors acting upstream of the pathway. This shows that pathways were not formed/refined in the same way, nor according to the same timing during evolution.

This study could have been completed by an analysis based on known pathways in yeast to validate or invalidate certain interactions that are considered to exist if the two genes are present. However, only one pathway from our list, MAKP, is described in KEGG for the *Saccharomyces cerevisiae* species, this pathway being widely documented (Chavel et al., 2014; Saito, 2010; Zou et al., 2008), including in plants (Meng & Zhang, 2013). In the literature, other pathways are also documented in yeast, such as cAMP (Portela & Rossi, 2020; Tamaki, 2007), mTOR (Powers et al., 2004), Ras (Tisi et al., 2014) or Sphingolipid (Montefusco et al., 2014). Thus, using the KEGG database, it is not possible to establish an exhaustive evolutionary bridge of signaling pathways between yeast and humans. Moreover, some species lack certain pathways, such as yeast, which lacks the NF-kappa B pathway (Ho et al., 2017; Saleski et al., 2017), and yet some elements of the pathway are present in *Saccharomyces cerevisiae* such as caseins kinase 2 (CSNK2A1/2/3 and CSNK2B). In yeast, these caseins are known to be essential for mitophagy (Kanki et al., 2013).

Another point of discussion concerns the methodology of the dating of the appearance of a gene. For this we are limited by the different versions of Genomicus and Genomicus Metazoa. In particular, the "intersection species" between these two databases are *Drosophila melanogaster* and *Caenorhabditis elegans*. Indeed, if for a given gene the oldest ortholog found is not one of these two species, or if this gene is lost in both species but present in more ancient taxa, our methodology does not allow us to use

adequately the Genomicus Metazoa trees. Genomicus trees are modified trees of Ensembl, and therefore the limit comes from its origin Ensembl. Furthermore, concerning Ensembl, depending on the versions, there may be mega-trees with all the paralogs of a gene in the same tree, or in sub-trees, with one paralog per sub-tree. This change from mega-trees into several sub-trees in Ensembl V94 and further (Emily, 2018) –which is not automatic, it depends on the size of the paralog family–, makes the attribution of appearance times more complex.

Among the 25 clades that we have selected, not all have been studied to the same extent. As shown in Suppl Data 4, we can see that Yeast (clade 1) only includes one species (*Saccharomyces cerevisiae*) while for Aves (clade 19), we included 13 species. However, our data show that appearance of the relevant genes is concentrated on 2 nodes (those subtending Opisthokonta, and Vertebrates) and the genes that appeared on these branches represent 75.7% of all the genes involved in the KEGG pathways.

It must also be considered that the signaling pathways noted on KEGG (but more broadly on databases) are human representations, and simplifications have been made to simplify reading and understanding. However, as seen in Figure 8 A, some proteins are involved in multiple pathways, such as the Shc → Grb2 → SOS → Ras → Raf1 → MEK → ERK subpathway which is involved (partially or entirely) in 22 pathways (ERBB, Estrogen, GnRH, Insulin, Prolactin, Relaxin, B cell receptor, Chemokine, F epsilon RI, T cell receptor, Neurotrophin, cAMP, FoxO, JAK-STAT, MAKP, mTOR, Phospholipase D, PI3K-Akt, Rap1, Ras, Sphingolipid, VEGF). The RTK/RAS/ERK component is indeed known to be common to *Drosophila*, nematode and humans (Ashton-Beaucage & Therrien, 2010).

## Conclusion

This study highlights two key phases of *Opisthokonta* evolution concerning the genes encoding proteins involved in signaling pathways in animals, near the base of *Opisthokonta* and of Vertebrata. We also observe a correlation (positive or negative depending of the pathway) between the position of the proteins in each pathway and the node of birth of their corresponding gene in the tree of life.

# Material and methods

## Implementation of the database

### *Signaling pathways by KEGG*

We followed the methodology described in our precedent paper (Picolo et al., in press).

We use the KEGG pathways because they are annotated in humans and therefore the genes are annotated in humans, and we use parsimony and the trees to locate the branch on which each gene originated. We retrieved a list of 2,298 unique genes encoding proteins involved in the 47 human intracellular signaling pathways available on the KEGG V104.0 database (https://www.genome.jp/kegg) (Bader et al., 2006) using the keywords “signaling pathway” and “human” (Table II). KEGG is one of the most referenced and used databases listing signaling pathways.

Each signaling pathway was retrieved in .xml (Extensible Markup Language) format from KEGG's PATHWAY Database tool. This file format is well supported by R libraries such as XML (Lang & Kalibera, 2023) and igraph (Csárdi et al., 2023; Csárdi & Nepusz, 2006).

The gene products of KEGG pathways are inscribed in rectangular blocks that we call "labels", for the XML file, this is the label "name", and for simplicity for readers, these labels can cover several paralogs (Figure 9). Some paralogous genes appear under different labels, as with the PPAR pathway for which the different proteins PPARα, PPARγ and PPARδ have similar interactions (with FABP3, Thiolase B, aP2 and UCP-1) and specific interactions (PPARα → HMGCS2; PPARγ → GyK for example). Each pathway is constructed in such a way that there can be one or more inputs (e.g., a ligand) and one or more outputs (e.g., transcription factor). Each sub-pathway is analyzed, which means that the same gene can be involved in several sub-pathways (e.g.: A → B → C and A → D → E).

### *Tree of life from Genomicus*

To determine the time of appearance, two phylogenetic trees from Genomicus (Nguyen et al., 2018) were used: the vertebrate tree and the metazoan tree. The vertebrate tree is the V109 tree (https://www.genomicus.bio.ens.psl.eu/genomicus-109.01), comprising 199 animal species and covering a range of species belonging to the vertebrate clade, as well as *Drosophila melanogaster* and *Caenorhabditis elegans* and to this is added a yeast : *Saccharomyces Cerevisiae*. The metazoan tree is that of V51 (https://www.genomicus.bio.ens.psl.eu/genomicus-metazoa-51.01), comprises 116 animal species and covers all main taxa belonging to the non-vertebrate metazoan clade. *Drosophila melanogaster* and *Caenorhabditis elegans* are the two “intersection species” between these two databases. (Figure 1, Suppl. Data 4).

From these Genomicus trees, we determined 25 clades on a similar and “enriched” model of our previous work (Grandchamp & Monget, 2018): 1 : Yeast (~ 1110 my), 2 : Porifera (~ 725 my), 3 : Placozoa (~ 550 my), 4 : Ctenophora (~ 109 my), 5 : Cnidaria (~ 588 my), 6 : Xenacoelomorpha (~ 550 my), 7 :

Spiralia (~ 610 my), 8 : Ecdysozoa (~ 653 my), 9: Echinodermata (~ 596 my), 10 : Cephalochordata (520 my), 11: Tunicata (~ 446 my), 12 : Petromyzontidae (~ 416 my), 13: Myxini (~ 360 my), 14: Chondrichtyes (~ 413 my), 15: Actinopterygii (~ 396 my), 16: Coelacanthidae (~ 350 my), 17: Amphibia (~ 319 my), 18: Testudines (~ 194 my), 19: Aves (~ 109 my), 20: Crocodylia (~ 87 my), 21: Squamata (~ 189 my), 22: Sphenodontia (~ 200 my), 23: Prototheria (~ 166 my), 24: Eutheria (~ 98 my), 25: Metatheria (~ 66 my). To provide consistent ages, and given that several of the oldest nodes are poorly constrained by the fossil record, we have used molecular time estimates throughout (de Vienne, 2016; Kumar et al., 2022; Nguyen et al., 2021). These clades are our references for the present study. The gene is considered to appear on the branch that leads to the clade for which all included taxa possess the same gene.

## Dating the appearance of genes

We consider the node of appearance by the presence of an ortholog of our gene within clades (e.g.: gene A has an ortholog shared by humans and *Saccharomyces cerevisiae*, so its appearance is assumed to predate Opisthokonta) (https://github.com/florianepicolo/birth-gene).

Some cases require a more complex analysis. If under a label in the KEGG pathways, several paralogues are listed, then the latest possible date of origin of this gene is the oldest node at which paralogs occur. If several paralogs are present in a pathway but under different labels, we considered them independently from other paralogs. In the case of a protein complex, we considered the complex to be a group, so we considered the node of birth of the complex the deepest node for which all the genes of the complex are present, considering that the complex cannot be functional unless all the proteins are present.

All elements in a pathway are assigned a rank in the pathway, and multiple ranks are possible for the elements that are present in several sub-pathways. We have not considered the “components” of the pathways present on KEGG in the allocation of the ranks if they are not proteins, and therefore do not involve genes, like elements like Ca2+ or lactate (example: $A_1$ → $B_2$ → Ca2+ → $C_3$, with ranks in index).

Furthermore, in an interaction A → B (A is closer to the receptor than B), the interaction is said to be “forward” if gene A was born after B, “backward” if gene A was born before B, and “simultaneous” if A and B were born on the same branch.

In the present study, ten pathways are involved in immunity (Table II), their genes appearing later (from the vertebrates) in the tree of life than the average (chi test p = 3.50E-08)

## Analyses and statistics

To determine whether there is more forward, backward, or simultaneous appearance interaction compared to random interactions, we performed permutation test (1000 permutations). We performed

correlation test (Pearson) to determine a possible relation between the branch of appearance of a gene and the position (upstream or downstream) of the corresponding protein in the pathway.

The distribution of relationships also needs to be considered. Indeed, the relationships of the same types (i.e. backward, forward, and simultaneous) can be spread in the pathways or clustered. To answer whether these clusters are bigger or smaller compared to randomness, we adopted a simulation approach described hereafter and for which a visual flowchart is available in sup. Mat X. For each pathway, we extracted the components of the five different types: forward (target younger than origin), forward and simultaneous (target's age equal or lower than origin), simultaneous (equal age), backward (target older than origin), backward and simultaneous (target's age equal or higher than origin). For each five types of relationships, we extracted the distribution of these clusters sizes. We then simulated 1000 alternative pathways under the null hypothesis (H0) that the relationships between genes have no relationship with their age. We did that simply by permuting relationships between genes without altering ages. For each simulated pathway under H0, we obtained five corresponding distributions of clusters sizes. To verify if these latter have bigger or smaller clusters sizes, we did a permutation test based on means. We finally considered how frequent random pathways have smaller or bigger clusters sizes, how much is the difference between these distributions (raw vs simulated under H0) and how frequent it is significant.

**Table I - Correlation between time of birth and rank in the lane**

| **Pathway** | **Number of subpath** | **Max subpath** | **Birth min, max** | **Correlation (r)** | **pvalue** |
|---|---|---|---|---|---|
| Adipocytokine | 54 | 8 | [1,17] | **-0.58** | *2.44E-24** |
| Fc epsilon RI | 42 | 10 | [1,24] | **-0.53** | *2.02E-16** |
| VEGF | 19 | 10 | [1,18] | **-0.53** | *3.15E-08** |
| PPAR | 61 | 2 | [1,23] | -0.48 | *4.41E-08** |
| Insulin | 255 | 11 | [1,25] | -0.42 | *1.28E-78** |
| Rap1 | 108 | 5 | [1,23] | -0.39 | *2.01E-16** |
| B cell receptor | 88 | 13 | [1,22] | -0.38 | *2.76E-19** |
| C-type lectin receptor | 107 | 8 | [1,25] | -0.36 | *7.24E-13** |
| Thyroid hormone | 102 | 7 | [1,21] | -0.34 | *2.11E-11** |
| Sphingolipid | 51 | 6 | [1,23] | -0.32 | *1.26E-06** |

| | | | | | |
|---|---|---|---|---|---|
| Estrogen | 29 | 16 | [2,24] | -0.31 | *1.14E-04** |
| Apelin | 26 | 11 | [1,21] | -0.31 | *4.36E-06** |
| TNF | 27 | 8 | [1,24] | -0.29 | *4.15E-04** |
| ErbB | 309 | 10 | [1,19] | -0.26 | *6.67E-31** |
| cGMP-PKG | 201 | 12 | [1,25] | -0.23 | *2.76E-16** |
| Glucagon | 40 | 9 | [1,25] | -0.22 | *2.92E-03** |
| Oxytocin | 51 | 9 | [1,21] | -0.22 | *1.88E-04** |
| mTOR | 375 | 14 | [1,16] | -0.21 | *9.75E-31** |
| HIF-1 | 54 | 7 | [1,17] | -0.20 | *1.38E-02** |
| GnRH | 16 | 16 | [1,21] | -0.19 | *2.00E-02** |
| MAPK | 2083 | 11 | [1,21] | -0.18 | *3.73E-104** |
| Phospholipase D | 235 | 11 | [2,25] | -0.18 | *3.61E-14** |
| TGF-beta | 61 | 7 | [1,20] | -0.15 | *2.05E-02** |
| Neurotrophin | 952 | 16 | [1,25] | -0.13 | *8.28E-35** |
| Chemokine | 183 | 11 | [1,21] | -0.10 | *1.52E-04** |
| Ras | 232 | 9 | [1,23] | 0.06 | *3.46E-02** |
| NOD-like receptor | 420 | 9 | [1,25] | 0.06 | *7.44E-04** |
| cAMP | 1966 | 11 | [1,24] | 0.07 | *6.55E-14** |
| IL-17 | 709 | 9 | [1,25] | 0.07 | *2.12E-05** |
| T cell receptor | 376 | 9 | [1,25] | 0.08 | *2.33E-04** |
| JAK-STAT | 12958 | 12 | [1,25] | 0.11 | *2.79E-280** |
| NF-kappa B | 110 | 6 | [1,24] | 0.17 | *1.55E-04** |
| Wnt | 402 | 11 | [1,24] | 0.22 | *6.55E-29** |
| Calcium | 77 | 8 | [1,21] | 0.36 | *1.16E-09** |
| RIG-I-like receptor | 374 | 8 | [1,19] | 0.49 | *3.01E-142** |

| | | | | | |
|---|---|---|---|---|---|
| Ovarian steroidogenesis | 15 | 6 | [1,25] | -0.22 | *6.13E-02** |
| FoxO | 70 | 9 | [1,25] | -0.10 | *8.54E-02** |
| PI3K-Akt | 668 | 13 | [1,25] | -0.02 | 1.62E-01 |
| p53 | 254 | 7 | [1,25] | 0.04 | 1.76E-01 |
| AMPK | 184 | 10 | [1,25] | 0.04 | 2.70E-01 |
| Toll-like receptor | 371 | 10 | [1,25] | 0.02 | 2.97E-01 |
| AGE-RAGE | 358 | 6 | [1,24] | -0.02 | 4.53E-01 |
| Hedgehog | 41 | 5 | [1,24] | -0.05 | 6.23E-01 |
| Relaxin | 140 | 13 | [1,24] | 0.01 | 6.28E-01 |
| Notch | 64 | 4 | [1,20] | 0.02 | 7.66E-01 |
| Prolactin | 63 | 14 | [1,25] | -0.01 | 7.96E-01 |
| Hippo | 73 | 6 | [1,24] | 0.01 | 9.03E-01 |

The correlation index (r) quantifies the strength of the relationship, it is said to be strong when it is between [-1, -0.5] and [0.5, 1] (in bold on Table 1), and it is said to be low between [-0.5, 0[ and ]0, 0.5], and without correlation to r = 0. Significant values (*, $p<0.05$) are written in italics.

**Table II - List of signaling pathways and their characteristics**

| ***Signaling pathway*** | ***KEGG categories*** | **Number of genes** | ***Number of interactions*** |
|---|---|---|---|
| p53 | Cell growth and death | 64 | 70 |
| AGE-RAGE | Endocrine and metabolic disease | 62 | 93 |
| Adipocytokine | Endocrine system | 36 | 48 |
| Estrogen | Endocrine system | 62 | 69 |
| Glucagon | Endocrine system | 50 | 55 |
| GnRH | Endocrine system | 41 | 39 |
| Insulin | Endocrine system | 62 | 77 |
| Ovarian steroidogenesis | Endocrine system | 45 | 25 |
| Oxytocin | Endocrine system | 58 | 73 |
| PPAR | Endocrine system | 51 | 63 |
| Prolactin | Endocrine system | 54 | 55 |
| Relaxin | Endocrine system | 81 | 103 |
| Thyroid hormone | Endocrine system | 78 | 85 |
| B cell receptor | Immune system | 47 | 57 |
| C-type lectin receptor | Immune system | 154 | 185 |
| Chemokine | Immune system | 58 | 82 |
| FC epsilon RI | Immune system | 42 | 47 |
| IL-17 | Immune system | 91 | 152 |
| NOD-like receptor | Immune system | 168 | 164 |
| RIG-I-like receptor | Immune system | 53 | 73 |
| T cell receptor | Immune system | 66 | 99 |
| Toll-like receptor | Immune system | 79 | 109 |
| Neurotrophin | Nervous system | 77 | 117 |
| AMPK | Signal transduction | 69 | 67 |
| Apelin | Signal transduction | 62 | 78 |
| Calcium | Signal transduction | 52 | 67 |
| cAMP | Signal transduction | 88 | 120 |
| cGMP-PKG | Signal transduction | 65 | 74 |
| ErbB | Signal transduction | 60 | 91 |
| FoxO | Signal transduction | 80 | 78 |
| Hedgehog | Signal transduction | 59 | 50 |
| HIF-1 | Signal transduction | 65 | 76 |
| Hippo | Signal transduction | 91 | 85 |
| JAK-STAT | Signal transduction | 85 | 262 |
| MAPK | Signal transduction | 119 | 172 |
| mTOR | Signal transduction | 76 | 91 |

| | | | |
|---|---|---|---|
| NF-Kappa B | Signal transduction | 137 | 122 |
| Notch | Signal transduction | 25 | 30 |
| Phospholipase D | Signal transduction | 56 | 71 |
| PI3K-Akt | Signal transduction | 90 | 97 |
| Rap1 | Signal transduction | 80 | 99 |
| Ras | Signal transduction | 87 | 112 |
| Sphingolipid | Signal transduction | 63 | 72 |
| TGF-Beta | Signal transduction | 73 | 76 |
| TNF | Signal transduction | 101 | 54 |
| VEGF | Signal transduction | 28 | 34 |
| Wnt | Signal transduction | 85 | 98 |

## Legends figures

**Figure 1 – Simplified animal tree of life and clades of study**

Each rectangle represents a node of speciation during evolution. Each branch represents a terminal clade of the tree with its associated number in the colored circles. The length of the branches is not representative of the actual evolutionary divergence. In bold are the clades available on Genomicus Vertebrates, and in italics are the clades available on Genomicus Metazoa.

**Figure 2 - Schematic representation of three possible scenarios linking the order of appearance of a gene and the position (rank) of the corresponding protein in the pathway**

A–C represent the three possible scenarios. The colors represent branches of birth for a gene (from 1 to 25). We arbitrarily chose ten proteins/positions per pathway for the illustration. In scenario A, the order of proteins in the pathway matches the order of appearance of genes in the tree. Scenario B is characterized by the opposite relationship (reverse order between protein position in the pathway and of the genes coding these proteins). In the scenario C, there is no link between the position of the protein in the pathway and the order of appearance of the corresponding gene.

**Figure 3 - Distribution of birth of genes encoding proteins involved in the 47 signaling pathways on the branches subtending 25 nodes**

Distribution of nodes of birth for each of the genes encoding proteins involved in a signaling pathway. In the text, these are sometimes referred to by number (1 being Opisthokonta, not shown here), which matches the order shown here (i.e. Methateria is node 25).

**Figure 4 – Backward, forward or simultaneous interaction for each signaling pathway**

Each horizontal column represents a signaling pathway (described on the left). On each column, for a A → B interaction, the orange color corresponds to simultaneous interaction (A was born at the same node

as B), the light green color corresponds to forward interaction (A was born after B), the green color corresponds to backward interaction (A was born before B).

**Figure 5 - Distribution of differences (delta) in the node of birth of genes encoding two partners for each pathway**

Boxplot distribution for each pathway of the difference (delta) between the positions (as rank order, disregarding absolute age) of the node of birth of a gene encoding one protein and of the node of birth encoding one of its partner(s). The pathways are centered on their median (dark vertical line in the center of the box), representing the middle half of the data between the first quartile (Q1) and the third quartile (Q3). Horizontal lines extend from the box to show data dispersion of the data, and isolated points indicates outliers.

**Figure 6 - Distribution of deltas of node of birth of genes encoding proteins involved in the Hippo pathway, according to the node of birth of each gene**

Deltas are calculated via number of clade of birth for the A gene - clade of birth for the B gene for the interaction A → B. For example, if gene A was born on the branch below the blue clade (clade 1), and its delta with B is -10, then clade of birth of B is 11. Moreover, in this case, it is a backward relationship, because A was born before B. The distributions for all the other pathways are shown in (Suppl. Data 1).

**Figure 7 – Node of birth of genes encoding proteins involved in each pathway according of the position (rank) of the protein in the pathway**

Each horizontal column corresponds to one pathway. On the left, the upstream proteins of the pathway (close to the cell membrane), on the right, the downstream proteins of the pathway (close to the nucleus). Depending on signaling pathways, there are between two to sixteen ranks on the abscissa. Each rectangle represents for a given rank the distribution of nodes of birth for each of the proteins occupying the rank. For example, for the GnRH pathway at position/rank 1, all the proteins have appeared at clade 14 (*Euteleostomi*). The pathways are sorted by correlation index in Table I. The distribution of positions/ranks and births of the different proteins for each pathway is accessible in (Suppl. Data 2).

**Figure 8 – Examples of colored KEGG pathways depending on the node of birth of each protein**

Each color represents a clade. The white rectangles correspond to the genes for which we have not been able to determine the node of birth due to lack of information about the gene. The KEGG legend is available here: https://www.genome.jp/kegg/document/help_pathway.html. (A) Hippo signaling pathway; (B) PPAR signaling pathway; (C) Notch signaling pathway. All colored pathways are available in (Suppl. Data 3).

**Figure 9 - Different labels in the KEGG database (example of the Notch pathway)**

A: Graphic representation of the Notch pathway on the KEGG web page. Each rectangle represents one or more proteins of the pathway; the arrows represent an activation, whereas lines that do not end with an arrow but with another short perpendicular represent an inhibition, expression...). B: Extract from the Notch pathway .xml file. For this example, Fringe, which corresponds to the FGN gene corresponds to "id 33" in the .xml file, and we note that there are 3 hsa identifiers, i.e., 3 paralogs: hsa:3955 (LFNG), hsa:4242 (MFNG) and hsa:5986 (RFNG). Next, PSE2, which corresponds to the PSENEN gene, corresponds to "id 26" that has no other paralog involved in this interaction.

## Supplementary data

All additional data are sorted in the following order: p53, AGE-RAGE, Adipocytokine, Estrogen, Glucagon, GnRH, Insulin, Ovarian steroidogenesis, Oxytocin, PPAR, Prolactin, Relaxin, Thyroid hormone, B cell receptor, C-type lectin receptor, Chemokine, FC epsilon RI, IL-17, NOD-like receptor, RIG-I-like receptor, T cell receptor, Toll-like receptor, Neurotrophin, AMPK, Apelin, Calcium, cAMP, cGMP-PKG, ErbB, FoxO, Hedgehog, HIF-1, Hippo, JAK-STAT, MAPK, mTOR, NF-Kappa B, Notch, Phospholipase D, PI3K-Akt, Rap1, Ras, Sphingolipid, TGF-Beta, TNF , VEGF, Wnt.

**Suppl. Data 1 - Distribution of deltas of node of birth of genes encoding proteins involved in all the pathways, according to the node of birth of each gene**

Deltas are calculated via clade of birth rank for the A gene - clade of birth rank for the B gene for the interaction A → B. For example, if gene A was born at the blue clade (clade 1), and the clade of birth of B is 11, the A → B delta is -10. Moreover, in this case, it is a backward relationship, because A was born before B.

**Suppl. Data 2 - Distribution of genes by position/rank in the pathway and node of birth**

Each graph represents one of the 47 pathways. Abscissa: the different proteins involved in the pathway; ordinate: position/rank of the protein within the pathway. Each protein is colored depending on the node of birth of its corresponding gene. Proteins are represented by a dot, are characterized by their position(s) they occupy within the pathway. The size of the dots is proportional to the number of times they are in these positions.

**Suppl. Data 3 - Colored KEGG pathways depending on the node of birth of each protein**

Each color represents a clade. The white rectangles correspond to the genes for which we have not been able to determine the node of birth due to lack of information about the gene. The KEGG legend is available here: https://www.genome.jp/kegg/document/help_pathway.html.

**Suppl. Data 4 - Animal tree of life and clades of study**

Tree of life of the 315 species studied here, generated using the information available in Ensembl and Ensembl Metazoa, and with R's ape package (Paradis & Schliep, 2019). The tree is rooted to reflect the phylogeny of the *Opisthokonta*, and the branches are not to scale. The colors are those used in the figures of the article. Each clade is represented by one or more species in our database.

## References


Albelda, S. M., & Buck, C. A. (1990). Integrins and other cell adhesion molecules. *FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology*, *4*(11), 2868-2880.

Andreani, J., & Guerois, R. (2014). Evolution of protein interactions : From interactomes to interfaces. *Archives of Biochemistry and Biophysics*, *554*, 65-75. https://doi.org/10.1016/j.abb.2014.05.010

Ashton-Beaucage, D., & Therrien, M. (2010). La signalisation RTK/RAS/ERK élargie—Contributions de la génétique à l'assemblage d'un réseau de signalisation. *médecine/sciences*, *26*(12), Article 12. https://doi.org/10.1051/medsci/201026121067

Bader, G. D., Cary, M. P., & Sander, C. (2006). Pathguide : A pathway resource list. *Nucleic Acids Research*, *34*(Database issue), D504-506. https://doi.org/10.1093/nar/gkj126

Benoit, G., Cooney, A., Giguere, V., Ingraham, H., Lazar, M., Muscat, G., Perlmann, T., Renaud, J.-P., Schwabe, J., Sladek, F., Tsai, M.-J., & Laudet, V. (2006). International Union of Pharmacology. LXVI. Orphan nuclear receptors. *Pharmacological Reviews*, *58*(4), 798-836. https://doi.org/10.1124/pr.58.4.10

Berger, M., Stahl, N., Del Sal, G., & Haupt, Y. (2005). Mutations in proline 82 of p53 impair its activation by Pin1 and Chk2 in response to DNA damage. *Molecular and Cellular Biology*, *25*(13), 5380-5388. https://doi.org/10.1128/MCB.25.13.5380-5388.2005

Bloom, J. D., Wilke, C. O., Arnold, F. H., & Adami, C. (2004). Stability and the evolvability of function in a model protein. *Biophysical Journal*, *86*(5), 2758-2764. https://doi.org/10.1016/S0006-3495(04)74329-5

Csárdi, G., & Nepusz, T. (2006). *The igraph software package for complex network research*. https://www.semanticscholar.org/paper/The-igraph-software-package-for-complex-network-Cs%C3%A1rdi-Nepusz/1d2744b83519657f5f2610698a8ddd177ced4f5c

Csárdi, G., Nepusz, T., Müller, K., Horvát, S., Traag, V., Zanini, F., & Noom, D. (2023). *igraph for R : R interface of the igraph library for graph theory and network analysis* [Logiciel]. Zenodo. https://doi.org/10.5281/zenodo.8046777

de Vienne, D. M. (2016). Lifemap : Exploring the Entire Tree of Life. *PLOS Biology*, *14*(12), e2001624. https://doi.org/10.1371/journal.pbio.2001624

Echave, J., & Wilke, C. O. (2016). *Biophysical models of protein evolution : Understanding the patterns of evolutionary sequence divergence* (p. 072223). bioRxiv. https://doi.org/10.1101/072223

Fraser, H. B., Hirsh, A. E., Wall, D. P., & Eisen, M. B. (2004). Coevolution of gene expression among interacting proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(24), 9033-9038. https://doi.org/10.1073/pnas.0402591101

Grandchamp, A., & Monget, P. (2018). Synchronous birth is a dominant pattern in receptor-ligand evolution. *BMC Genomics*, *19*. https://doi.org/10.1186/s12864-018-4977-2

Grandchamp, A., & Monget, P. (2020). The membrane receptors that appeared before their ligand : The different proposed scenarios. *PloS One*, *15*(5), e0231813. https://doi.org/10.1371/journal.pone.0231813

Howard, A. D., McAllister, G., Feighner, S. D., Liu, Q., Nargund, R. P., Van der Ploeg, L. H., & Patchett, A. A. (2001). Orphan G-protein-coupled receptors and natural ligand discovery. *Trends in Pharmacological Sciences*, *22*(3), 132-140. https://doi.org/10.1016/s0165-6147(00)01636-9

Jack, B. R., Meyer, A. G., Echave, J., & Wilke, C. O. (2016). Functional Sites Induce Long-Range Evolutionary Constraints in Enzymes. *PLoS Biology*, *14*(5), e1002452. https://doi.org/10.1371/journal.pbio.1002452

Jones, S. N., Roe, A. E., Donehower, L. A., & Bradley, A. (1995). Rescue of embryonic lethality in Mdm2-deficient mice by absence of p53. *Nature*, *378*(6553), 206-208. https://doi.org/10.1038/378206a0

Kachroo, A. H., Laurent, J. M., Yellman, C. M., Meyer, A. G., Wilke, C. O., & Marcotte, E. M. (2015). Evolution. Systematic humanization of yeast genes reveals conserved functions and genetic modularity. *Science (New York, N.Y.)*, *348*(6237), 921-925. https://doi.org/10.1126/science.aaa0769

Kanki, T., Kurihara, Y., Jin, X., Goda, T., Ono, Y., Aihara, M., Hirota, Y., Saigusa, T., Aoki, Y., Uchiumi, T., & Kang, D. (2013). Casein kinase 2 is essential for mitophagy. *EMBO Reports*, *14*(9), 788-794. https://doi.org/10.1038/embor.2013.114

Kauffman, S. A. (1993). *The Origins of Order : Self-organization and Selection in Evolution*. Oxford University Press.

Kent, R. S., De Lean, A., & Lefkowitz, R. J. (1980). A quantitative analysis of beta-adrenergic receptor interactions : Resolution of high and low affinity states of the receptor by computer modeling of ligand binding data. *Molecular Pharmacology*, *17*(1), 14-23.

Khoury, C. M., & Greenwood, M. T. (2008). The pleiotropic effects of heterologous Bax expression in yeast. *Biochimica Et Biophysica Acta*, *1783*(7), 1449-1465. https://doi.org/10.1016/j.bbamcr.2007.12.013

Koshland, D. E. (1958). Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, *44*(2), 98-104. https://doi.org/10.1073/pnas.44.2.98

Kumar, S., Suleski, M., Craig, J. M., Kasprowicz, A. E., Sanderford, M., Li, M., Stecher, G., & Hedges, S. B. (2022). TimeTree 5 : An Expanded Resource for Species Divergence Times. *Molecular Biology and Evolution*, *39*(8), msac174. https://doi.org/10.1093/molbev/msac174

Lang, D. T., & Kalibera, T. (2023). *XML : Tools for Parsing and Generating XML Within R and S-Plus* (3.99-0.14) [Logiciel]. https://cran.r-project.org/web/packages/XML/index.html

Lewis, A. C. F., Saeed, R., & Deane, C. M. (2010). Predicting protein-protein interactions in the context of protein evolution. *Molecular bioSystems*, *6*(1), 55-64. https://doi.org/10.1039/b916371a

Lovell, S. C., & Robertson, D. L. (2010). An integrated view of molecular coevolution in protein-protein interactions. *Molecular Biology and Evolution*, *27*(11), 2567-2575. https://doi.org/10.1093/molbev/msq144

Lynch, M., & Hagner, K. (2015). Evolutionary meandering of intermolecular interactions along the drift barrier. *Proceedings of the National Academy of Sciences*, *112*(1). https://doi.org/10.1073/pnas.1421641112

Markov, G. V., Gutierrez-Mazariegos, J., Pitrat, D., Billas, I. M. L., Bonneton, F., Moras, D., Hasserodt, J., Lecointre, G., & Laudet, V. (2017). Origin of an ancient hormone/receptor couple revealed by resurrection of an ancestral estrogen. *Science Advances*, *3*(3), e1601778. https://doi.org/10.1126/sciadv.1601778

Maslov, S., & Sneppen, K. (2002). Specificity and stability in topology of protein networks. *Science (New York, N.Y.)*, *296*(5569), 910-913. https://doi.org/10.1126/science.1065103

Mintseris, J., & Weng, Z. (2005). Structure, function, and evolution of transient and obligate protein-protein interactions. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(31), 10930-10935. https://doi.org/10.1073/pnas.0502667102

Montes de Oca Luna, R., Wagner, D. S., & Lozano, G. (1995). Rescue of early embryonic lethality in mdm2-deficient mice by deletion of p53. *Nature*, *378*(6553), 203-206. https://doi.org/10.1038/378203a0

Nguyen, N. T. T., Vincens, P., Dufayard, J. F., Roest Crollius, H., & Louis, A. (2021). Genomicus in 2022 : Comparative tools for thousands of genomes and reconstructed ancestors. *Nucleic Acids Research*, gkab1091. https://doi.org/10.1093/nar/gkab1091

Nguyen, N. T. T., Vincens, P., Roest Crollius, H., & Louis, A. (2018). Genomicus 2018 : Karyotype evolutionary trees and on-the-fly synteny computing. *Nucleic Acids Research*, *46*(D1), D816-D822. https://doi.org/10.1093/nar/gkx1003

Nooren, I. M. A., & Thornton, J. M. (2003). Diversity of protein-protein interactions. *The EMBO Journal*, *22*(14), 3486-3492. https://doi.org/10.1093/emboj/cdg359

Paradis, E., & Schliep, K. (2019). ape 5.0 : An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526-528. https://doi.org/10.1093/bioinformatics/bty633

Rand, D. M., Haney, R. A., & Fry, A. J. (2004). Cytonuclear coevolution : The genomics of cooperation. *Trends in Ecology & Evolution*, *19*(12), 645-653. https://doi.org/10.1016/j.tree.2004.10.003

Sullivan, S. M., & Holyoak, T. (2008). Enzymes with lid-gated active sites must operate by an induced fit mechanism instead of conformational selection. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(37), 13829-13834. https://doi.org/10.1073/pnas.0805364105

Thornton, J. W. (2001). Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(10), 5671-5676. https://doi.org/10.1073/pnas.091553298

Tisi, R., Belotti, F., & Martegani, E. (2014). Yeast as a model for Ras signalling. *Methods in Molecular Biology (Clifton, N.J.)*, *1120*, 359-390. https://doi.org/10.1007/978-1-62703-791-4_23

Williams, P. D., Pollock, D. D., & Goldstein, R. A. (2001). Evolution of functionality in lattice proteins. *Journal of Molecular Graphics & Modelling*, *19*(1), 150-156. https://doi.org/10.1016/s1093-3263(00)00125-x

Wuchty, S., Oltvai, Z. N., & Barabási, A.-L. (2003). Evolutionary conservation of motif constituents in the yeast protein interaction network. *Nature Genetics*, *35*(2), 176-179. https://doi.org/10.1038/ng1242

Zha, H., Fisk, H. A., Yaffe, M. P., Mahajan, N., Herman, B., & Reed, J. C. (1996). Structure-function comparisons of the proapoptotic protein Bax in yeast and mammalian cells. *Molecular and Cellular Biology*, *16*(11), 6494-6508. https://doi.org/10.1128/MCB.16.11.6494