The relationship between the hierarchical position of proteins in the human signal transduction network and their rate of evolution

David Alvarez-Ponce

Abstract

Background: Proteins evolve at disparate rates, as a result of the action of different types and strengths of evolutionary forces. An open question in evolutionary biology is what factors are responsible for this variability. In general, proteins whose function has a great impact on organisms’ fitness are expected to evolve under stronger selective pressures. In biosynthetic pathways, upstream genes usually evolve under higher levels of selective constraint than those acting at the downstream part, as a result of their higher hierarchical position. Similar observations have been made in transcriptional regulatory networks, whose upstream elements appear to be more essential and subject to selection. Less well understood is, however, how selective pressures distribute along signal transduction pathways.

Results: Here, I combine comparative genomics and directed protein interaction data to study the distribution of evolutionary forces across the human signal transduction network. Surprisingly, no evidence was found for higher levels of selective constraint at the upstream network genes (those occupying more hierarchical positions). On the contrary, purifying selection was found to act more strongly on genes acting at the downstream part of the network, which seems to be due to downstream genes being more highly and broadly expressed, performing certain functions and, in particular, encoding proteins that are more highly connected in the protein–protein interaction network. When the effect of these confounding factors is discounted, upstream and downstream genes evolve at similar rates. The trends found in the overall signaling network are exemplified by analysis of the distribution of purifying selection along the mammalian Ras signaling pathway, showing that upstream and downstream genes evolve at similar rates.

Conclusions: These results indicate that the upstream/downstream position of proteins in the signal transduction network has, in general, no direct effect on their rates of evolution, suggesting that upstream and downstream genes are similarly important for the function of the network. This implies that natural selection differently distributes across signal transduction networks and across biosynthetic and transcriptional regulatory networks, which might reflect fundamental differences in their function and organization.

Keywords: Network evolution, Evolutionary rate, Selective constraint, Purifying selection, Signal transduction, Molecular pathways, Upstream and downstream position
Background

Proteins’ rates of evolution vary across orders of magnitude (e.g., refs. [1,2]). Understanding the factors underlying this enormous variability is a central problem in evolutionary biology. The neutral theory of molecular evolution predicts higher selective pressures (and hence, slower rates of evolution) acting on genes encoding proteins that perform the most important functions for organisms’ biological fitness [3]; however, the relative contributions of genes to fitness remain elusive. Over the last decade, the availability of both genomic and functional data has allowed us to identify some of the factors that correlate with genes’ rates of evolution, among which patterns and levels of gene expression appear to be the most important (for review, see refs. [4-6]). However, this gene-centered approach has been able to explain only a relatively small fraction of the variability of rates of evolution (e.g., ref. [7]).

Genes and proteins do not act in isolation, but usually operate as components of complex networks of interacting molecules. Therefore, taking into account the interaction networks in which a gene participates may provide key insight into the evolutionary forces acting on it. Furthermore, understanding the function and evolution of molecular networks can aid applications such as metabolic engineering and drug discovery and design (for review, see refs. [8-10]). Molecular networks in general can be modeled as undirected graphs, whose nodes and edges represent proteins and interactions, respectively. Over the last years, the accumulation of interactomic data has allowed us to cast a first glance at the impact of genes’ position in the network on their patterns of molecular evolution (reviewed in refs. [11-13]). For instance, genes encoding the most connected proteins in metabolic and protein–protein interaction networks are more selectively constrained than those acting at the network periphery [14-19], and interacting genes tend to show correlated evolutionary histories, e.g. evolving at similar rates [14,16,20-24]. These observations indicate that the structure of molecular networks impose constraints on the evolution of their components.

In addition to the centrality of proteins in functional networks, their hierarchical position in molecular pathways has also been put forward as a determinant of their rates of evolution. Remarkably, in a number of biosynthetic pathways, the strength of purifying selection has been found to correlate with the position of genes along the upstream/downstream pathway axis, with upstream genes being generally the most constrained [25-30]. This distribution of the levels of selective constraint has been attributed to upstream genes being involved in the biosynthesis of a broader range of biochemical compounds than downstream genes, and hence being more pleiotropic [28], and/or to upstream genes exerting a higher influence over the flux of metabolites along pathways [31]. Consistent observations have been made in transcriptional regulatory networks: upstream genes in these networks are more likely to be essential [32], and simulation analysis has shown that genes exerting a higher degree of control over other genes, and/or those that are less regulated by other genes, are more strongly affected by selection [33].

Less well understood is how selective pressures distribute along signal transduction pathways. These pathways play a key role in modulating cell function in response to extracellular and intracellular stimuli. They are characterized by the presence of a receptor that, in response to signals, is able to trigger a cascade of events that ultimately modulate the final effectors, which are responsible for mediating the biological responses of the pathway. The signal is transmitted from the upstream to the downstream part of the pathway by means of a series of molecular interactions (e.g., phosphorylation and/or dephosphorylation events). These interactions have a directed nature, i.e., each interaction involves an upstream modulator (either activator or inhibitor) and a downstream modulated element (activated or inhibited). Therefore, signaling pathways and networks can be represented as directed graphs, in which nodes and arcs represent molecules and directed interactions, respectively. Despite the availability of some directed interaction networks (e.g., refs. [34-36]), their directed nature has often been overlooked in evolutionary analyses, being often treated as undirected networks.

So far, the distribution of levels of selective constraint across the upstream/downstream axis of signaling pathways has been studied in only a few pathways, with contrasting results. In the Drosophila Ras pathway, upstream genes appear to be more selectively constrained than those acting at the downstream part [37]. This is to be expected under a model in which upstream genes, owing to their higher hierarchical position, have a higher impact on the function of the pathway. Indeed, a single activated upstream protein has the potential to activate several downstream target molecules (e.g., ref. [38]); therefore, higher levels of selective constraint might be expected at these genes. However, the opposite pattern (i.e., stronger purifying selection acting on downstream genes) has been described in the Drosophila, Caenorhabditis and vertebrate insulin/TOR pathway [39-42], and in the yeast HOG pathway [43]. In both the yeast HOG pathway and the Caenorhabditis insulin/TOR pathway, the tendency appears to vanish when the effect of confounding factors that correlate with both network position and levels of selective constraint, such as expression level or codon bias, are removed [42,43], whereas in the Drosophila and vertebrate insulin/TOR pathway the polarity in the levels of selective constraint
is independent of these and other factors [39-41,44]. Therefore, the effect of upstream/downstream position of genes in signal transduction pathways on their rates of evolution, and the relative contribution of potential confounding factors, remain elusive.

Here, I combine comparative genomics and manually-curated directed protein interaction data to determine whether purifying selection differentially acts on the upstream and downstream genes of the human signal transduction network. For that purpose, new methods that explicitly take into account the directed character of the network are introduced. Surprisingly, no evidence was found supporting that upstream genes (those occupying higher hierarchical positions) are subject to increased levels of selective constraint. On the contrary, purifying selection was found to act more severely on genes acting at the downstream part of the network, which seems to be due to downstream genes being more highly and broadly expressed, having certain functions and, in particular, encoding proteins that are more highly connected in the protein–protein interaction network. When the effect of these confounding factors was discounted, upstream and downstream genes were found to evolve at similar rates. These results clearly indicate that the hierarchical position of genes in the signal transduction network has, in general, no direct effect on their rates of evolution, suggesting that genes occupying more hierarchical positions in the network are not more relevant for the function of the network and the fitness of the organism. The trend found in the overall signaling network is exemplified by analysis of the distribution of purifying selection along the mammalian Ras signaling pathway, showing that, contrary to previously suggested in Drosophila, upstream and downstream genes evolve at similar rates. These observations imply that natural selection differently distributes across signal transduction networks and across biosynthetic and transcriptional regulatory networks. This might reflect fundamental differences in the function and organization of the signal transduction network and metabolic and transcriptional regulatory networks.

**Results**

Unexpectedly, genes acting at the upstream part of the human signal transduction network exhibit higher rates of evolution

The human signal transduction network gleaned by Cui et al. [34] was used. In this network, directed edges (links) between pairs of proteins represent regulatory interactions (i.e., activating or inhibitory) — e.g., the directed edge “A → B” denotes that protein “A” activates or inhibits protein “B”. The dataset was assembled and manually curated by Cui et al. by merging the interactions compiled in different databases, followed by validation of each interaction by accurate revision of the literature. Undirected edges, which represent generic (i.e., non-signaling) physical interactions, were not considered in the current analysis, since they contain no information about the hierarchical (upstream/downstream) position of genes in the network. After filtering the dataset (see Methods), it comprised 1049 proteins connected by 2436 non-redundant directed interactions (activating or inhibitory). This dataset included 10 ligands, 102 receptors, 62 adapters, 187 kinases, 28 phosphatases, 123 transcription factors, 99 structural proteins, and 27 ion transmembrane transporters. Most edges represent direct physical protein–protein interactions, with the only exception of a small fraction of “transcription factor → target gene” edges.

For each gene in the network, the 1:1 mouse ortholog was identified and the impact of natural selection was characterized from the nonsynonymous to synonymous divergence ratio \( \omega = d_{\text{NS}}/d_{\text{S}} \). Values of \( \omega \) equal to 1 are expected for genes evolving neutrally, whereas \( \omega < 1 \) is indicative of the action of purifying selection preserving the sequence of the encoded proteins, and \( \omega > 1 \) in a number of codons is indicative of the action of positive selection (adaptive evolution) driving the fixation of nonsynonymous substitutions. The estimated \( \omega \) values range from 0.0001 to 0.814 (with a median value of 0.065), indicating that purifying selection has played an important role in the evolution of these genes.

I evaluated whether \( \omega \) values depend on the hierarchical position that genes occupy in the network. For that purpose, three different approaches were adopted to compare the evolutionary rates of upstream and downstream network genes. First, the \( \omega \) value of each gene was compared with the central value of the \( \omega \) values of all its direct downstream targets (the median in the case of an odd number, or one of the two central values in the case of an even number of downstream proteins; see Methods). Out of the 1049 genes in the network, 800 encode proteins that have at least one direct downstream target in the network (i.e., out-degree > 0). Of those, 446 genes exhibit a higher \( \omega \) value than their targets, whereas 352 show lower \( \omega \) values than their targets. This asymmetry is not compatible with a random distribution of \( \omega \) values across upstream and downstream genes (paired sign test, \( P = 0.001 \); Table 1). Therefore, surprisingly, upstream genes tend to show higher \( \omega \) values than the genes that they modulate. In order to establish whether the evolution of nonsynonymous or synonymous sites is responsible for this trend, this analysis was conducted for \( d_{\text{NS}} \) and \( d_{\text{S}} \) separately. Clearly significant differences were observed for \( d_{\text{NS}} \) (\( P = 3.50 \times 10^{-4} \)), but only marginally significant differences were observed for \( d_{\text{S}} \) (\( P = 0.040 \); Table 1), indicating that the evolutionary patterns of nonsynonymous sites are the main responsible for the trend observed in \( \omega \).
Next, evolutionary rates of genes encoding proteins occupying extreme upstream and downstream positions in the network were compared. A total of 333 genes were classified as upstream [as they modulate other proteins (out-degree > 0) but are not modulated by any other protein in the network (in-degree = 0)] and 249 were classified as downstream [modulated by at least one protein (in-degree > 0), but unable to modulate any other protein (out-degree = 0)]. Again, upstream genes were found to exhibit significantly higher \( \omega \) values: upstream and downstream genes exhibit median \( \omega \) values of 0.076 and 0.064, respectively (Mann–Whitney test, \( P = 0.037 \); Figure 1). Significant differences were also observed in the \( d_N \) (\( P = 0.049 \)), but not in the \( d_S \) (\( P = 0.782 \); Table 2) values.

Finally, I evaluated whether \( \omega \) values correlate with parameters that reflect the hierarchical position of genes in the network. Both in-degree (the number of upstream modulators or upcoming edges) and out-degree (number of downstream targets or outgoing edges) negatively correlate with \( \omega \) values, consistent with previous observations that highly connected genes are subject to stronger purifying selection [14-16]. Remarkably, the correlation is stronger for in-degree (Spearman’s rank correlation coefficient, \( \rho = -0.116, P = 1.74 \times 10^{-4} \)) than for out-degree (\( \rho = -0.070, P = 0.023 \)), indicating that the \( \omega \) value of a gene is more influenced by the number of upstream genes (i.e., by how regulated the encoded product is), than by the number of downstream targets (i.e., by the level of regulation that it exerts on the network). Similar results were obtained for \( d_N \), whereas \( d_S \) does not correlate with in-degree or out-degree (Table 3). Beyond considering the direct downstream targets of a gene, I also considered the effect of the total number of genes acting directly or indirectly downstream of a given gene (termed \( H \); see Methods) on its evolutionary rate. This parameter positively correlates with \( \omega \) (\( \rho = 0.067, P = 0.031 \)), again indicating that proteins potentially affecting the activity of a higher number of other proteins (i.e., those occupying higher hierarchical positions in the network) are encoded by genes that exhibit higher \( \omega \) values. Again, significant results were obtained for \( d_N \) (\( \rho = 0.067, P = 0.030 \)), but not for \( d_S \) (\( \rho = 0.016, P = 0.596 \); Table 3).

### Table 1 Paired tests comparing upstream genes with their downstream targets

| Parameter         | n \( ^a \) | Upstream > Downstream \( ^b \) | Upstream < Downstream \( ^c \) | \( P \)   |
|-------------------|-----------|-------------------------------|-------------------------------|--------|
| \( \omega \)      | 800       | 446                           | 352                           | 0.001*** |
| \( d_N \)         | 800       | 450                           | 348                           | 3.50\times10^{-4}*** |
| \( d_S \)         | 800       | 429                           | 370                           | 0.040*   |
| Expression level  | 763       | 357                           | 406                           | 0.082    |
| Expression breadth| 763       | 261                           | 337                           | 0.002**  |
| ENC               | 800       | 398                           | 402                           | 0.916    |
| Connectivity      | 709       | 230                           | 456                           | 8.66\times10^{-18}*** |
| Number of paralogs| 800       | 369                           | 373                           | 0.912    |

\* \( P < 0.05 \); ** \( P < 0.01 \); *** \( P < 0.001 \).

\( ^a \) Number of genes with at least one direct downstream target (out-degree>0). Only genes with available information for the parameter of interest were considered.

\( ^b \) Number of genes with a higher value than the central value of its downstream targets.

\( ^c \) Number of genes with a lower value than the central value of its downstream targets.

Figure 1 Comparison of genes encoding the upstream and downstream proteins of the human signal transduction network.
Table 2 Comparison of genes occupying extreme upstream and downstream positions in the signaling network

| Parameter         | Upstream |          |          | Downstream |          |          | P      |
|-------------------|----------|----------|----------|------------|----------|----------|--------|
|                   | n        | Median   | Average  | n          | Median   | Average  |        |
| ω                 | 333      | 0.076    | 0.112    | 249        | 0.064    | 0.098    | 0.037* |
| d_N               | 333      | 0.045    | 0.070    | 249        | 0.037    | 0.060    | 0.049* |
| d_S               | 333      | 0.561    | 0.630    | 249        | 0.567    | 0.626    | 0.782  |
| Expression level  | 322      | 27.32    | 116.28   | 243        | 32.43    | 165.13   | 0.239  |
| Expression breadth| 322      | 21.00    | 15.41    | 243        | 23.00    | 15.84    | 0.621  |
| ENC               | 333      | 50.43    | 49.36    | 249        | 50.36    | 49.39    | 0.712  |
| Connectivity      | 293      | 6.00     | 11.86    | 225        | 8.00     | 13.41    | 0.114  |
| Number of paralogs| 333      | 22.00    | 36.73    | 249        | 28.00    | 37.63    | 0.714  |

P-values were obtained from the Mann–Whitney U test. *, P < 0.05.

Taken together, these results show that evolutionary rates exhibit a polarity across the human signaling network, with downstream genes progressively exhibiting lower ω and d_N values. This observation is in sharp contrast with patterns observed in metabolic and transcriptional regulatory pathways, in which upstream genes are generally the ones exhibiting lower rates of evolution [25-30,32,33]. I next considered a number of potential factors that could account for the observed polarity.

The higher rates of evolution of upstream genes are the result of lower levels of purifying selection rather than a higher incidence of positive selection

The lower ω and d_N values observed in the downstream network genes suggest a higher strength of purifying selection acting on these genes. However, ω and d_N values are also affected by the action of positive selection, and hence the observed trends could alternatively be the result of a putatively higher incidence of positive selection on upstream genes. In order to discard this possibility, analyses were repeated after eliminating genes with the signature of positive selection.

When applied to pairs of human-mouse orthologs, the M7 vs. M8 test [45] identified 58 network genes with potential signatures of positive selection (i.e., a fraction of codons with ω > 1). Excluding these genes yielded similar results. First, after excluding these genes, 741 genes have at least one direct downstream target in the network, out of which 417 and 322 show higher and lower ω values than their targets, respectively (paired test, \( P = 0.001 \)). Significant results were also obtained for d_N \( (P = 3.79 \times 10^{-10}) \). Second, the 315 genes classified as upstream that display no signals of positive selection exhibit significantly higher ω (Mann–Whitney test, \( P = 0.033 \)), and d_N values \( (P = 0.049) \) than the 230 downstream genes without the signature of positive selection. Finally, ω and d_N values positively correlate with \( H \) \( (P = 0.065, P = 0.040) \) for ω; \( \rho = 0.064, P = 0.044 \) for d_N), and more strongly correlate with in-degree \( (\rho = -0.115, P = 2.71 \times 10^{-4}) \) for ω; \( \rho = -0.115, P = 2.94 \times 10^{-4} \) for d_N) than with out-degree \( (\rho = -0.070, P = 0.028 \) for ω; \( \rho = -0.072, P = 0.023 \) for d_N).

Because detection of positive selection is sensitive to the number of sequences used [46], analyses were...
repeated using a total of 8 mammalian genomes (human, chimpanzee, gorilla, orangutan, macaque, marmoset, mouse and rat). A total of 808 network genes exhibit a 1:1 orthology across all genomes, out of which 115 exhibit positive signals of positive selection. Again, excluding these genes did not alter the observations that genes acting at the downstream part of the network exhibit significantly lower $\omega$ (paired test, $P = 1.05 \times 10^{-3}$; Mann–Whitney test, $P = 0.033$; correlation with $H$, $\rho = 0.065$, $P = 0.040$) and $d_N$ values (paired test, $P = 0.001$; Mann–Whitney test, $P = 0.020$; correlation with $H$, $\rho = 0.072$, $P = 0.028$).

Therefore, the distribution of $\omega$ and $d_N$ values across the network may be the result of a polarity in the strength of purifying selection, with downstream genes being subject to higher levels of selective constraint, and not the result of a higher incidence of positive selection in the upstream part of the network.

Subcellular location does not account for the higher rates of evolution of upstream network proteins

Genes acting at the extracellular compartment have been reported to exhibit high rates of evolution [20,47]. Consistently, among genes in the used dataset, those acting at the extracellular space exhibit higher $\omega$ and $d_N$ values than the rest of the genes in the signal transduction network (Mann–Whitney test, $P = 1.56 \times 10^{-16}$ for $\omega$; $P = 2.23 \times 10^{-19}$ for $d_N$). This could potentially explain the observed polarity in $\omega$ and $d_N$ values if proteins occupying hierarchically positional interactions in the signaling network preferentially located to the extracellular space. However, I tested this possibility by comparing the proportion of extracellular proteins among proteins acting at the upstream (14.41%) and downstream (10.44%) parts of the network, finding no significant differences (Fisher’s exact test, $P = 0.168$). I.e., proteins occupying extreme upstream positions in the network do not tend to act at the extracellular space, thus making it unlikely that the observed high $\omega$ values in genes encoding the upstream part of the network would be linked to subcellular location. In order to further discard this possibility, analyses were repeated eliminating the 105 proteins targeted to the extracellular space, with similar results. First, genes are significantly less selectively constrained than their downstream direct targets (paired test, $P = 0.001$ for $\omega$; $P = 0.007$ for $d_N$). Second, the 285 non-extracellular proteins classified as upstream exhibit higher $\omega$ and $d_N$ values than the 223 non-extracellular proteins acting at the downstream part of the network, although the differences are not statistically significant (median $\omega$ for upstream and downstream genes: 0.068 and 0.059, respectively; Mann–Whitney test, $P = 0.110$; $P = 0.151$ for $d_N$). Third, the $\omega$ and $d_N$ values correlate better with in-degree ($\rho = -0.092$, $P = 0.005$ for $\omega$; $\rho = -0.091$, $P = 0.005$ for $d_N$) than with out-degree ($\rho = -0.051$, $P = 0.114$ for $\omega$; $\rho = -0.054$, $P = 0.100$ for $d_N$). Although evolutionary rates do not significantly correlate with $H$, the direction of the correlations is positive and comparable in strength to that observed in the overall dataset ($\rho = 0.052$, $P = 0.113$ for $\omega$; $\rho = 0.053$, $P = 0.102$ for $d_N$). These results indicate that the polarity observed in the levels of selective constraint is for the most part independent of the high evolutionary rates of proteins acting at the extracellular space.

Using an expanded dataset (which included also generic physical, undirected interactions), Cui, Purisima and Wang [20] found that signaling genes acting at the cell membrane, cytoplasm and nucleus exhibited different rates of evolution. Again, these differences raise the possibility that the observed polarity in the $\omega$ values across the network might be in part the result of the distribution of upstream and downstream proteins across the different subcellular compartments. However, I found that (1) genes acting at the different cell compartments do not exhibit significantly different rates of evolution for the subset of genes included in the current analysis (which contained only genes involved in directed signaling interactions) (median $\omega$ values of 0.052, 0.056 and 0.064, respectively; Kruskal–Wallis test, $P = 0.804$), and that (2) the upstream and downstream parts of the network do not contain different proportions of proteins locating to the cell membrane (Fisher’s exact test, $P = 0.738$), cytoplasm ($P = 0.773$) or nucleus ($P = 0.760$), making it unlikely that the polarity in the levels of selective constraint described here would obey to proteins’ subcellular location. In order to further discard this possibility, analyses were conducted separately on subsets of genes acting at the different cell locations, with consistent results across all four compartments. First, genes tend to evolve faster than their direct downstream targets acting at the same cell compartment, when the analyses are restricted to genes acting at the extracellular space (15 and 10 proteins exhibit higher and lower $\omega$ values, respectively, than their downstream targets acting in the same compartment; paired test, $P = 0.424$), the cell membrane (126 and 88 proteins; $P = 0.011$), the cytoplasm (186 vs. 177; $P = 0.675$), or the nucleus (142 vs. 138; $P = 0.858$). Second, $\omega$ values are higher for upstream than for downstream genes when genes acting at the extracellular space (median values of 0.164 and 0.158 for upstream and downstream genes, respectively), the cell membrane ($0.075 \text{ vs. } 0.063$), the cytoplasm (0.065 vs. 0.056) or the nucleus (0.0653 vs. 0.0646) are analyzed separately, although the differences are not statistically significant (Mann–Whitney test, $P > 0.05$). Taken together, these results indicate that the polarity in the levels of selective constraint observed across the network is independent of the subcellular location of its
components. The lack of significance in some tests probably owes to the reduced statistical power resulting from partitioning the dataset.

**Gene functions upstream and downstream of the signal transduction network**

Genes with different functions are known to evolve at different rates (e.g., refs. [23] and [48]). Hence, in order to gain further insight into the different selective pressures acting at the upstream and downstream parts of the human signaling network, it was tested whether the 333 genes acting at the upstream part performed different functions to the 249 genes acting at the downstream part. For that purpose, the frequencies of the GOSlim GOA terms of both groups was compared using the FatiGO software [49]. Upstream genes exhibited a significant enrichment in the term “kinase activity” ($q = 2.24 \times 10^{-4}$), whereas downstream genes were enriched in the category “ion transmembrane transporter activity” ($q = 2.65 \times 10^{-5}$). Hence, the higher levels of selective constraint observed in the downstream part of the signaling network could potentially result from this unequal distribution of gene functions, if genes encoding kinases evolved particularly fast, and/or those encoding ion transmembrane transporters evolved particularly slow. Both gene groups were found to evolve slower than the rest of genes in the network (median $\omega$ values for ion transporters, kinases, and the entire network: 0.039, 0.045, and 0.065, respectively; Mann–Whitney tests, $P = 3.99 \times 10^{-4}$ for ion transporters, $P = 2.39 \times 10^{-8}$ for kinases), raising the possibility that the higher levels of selective constraint acting at the downstream part of the network would be due to the enrichment of this part of the network in ion transmembrane transporters.

In order to discard this possibility, analyses were repeated excluding the 27 network genes encoding ion transmembrane transporters, with similar results, although not statistically significant in all cases. First, genes are significantly less selectively constrained than their downstream direct targets (paired test, $P = 0.003$ for $\omega$; $P = 2.62 \times 10^{-4}$ for $d_N$). Second, the 332 non-ion transporter proteins classified as upstream exhibit higher $\omega$ and $d_N$ values than the 230 non-ion transporter proteins acting at the downstream part of the network, although the differences are not statistically significant (median $\omega$ values for upstream and downstream genes: 0.068 and 0.076, respectively; Mann–Whitney test, $P = 0.181$ for $\omega$; $P = 0.203$ for $d_N$). Third, the $\omega$ and $d_N$ values correlate better with in-degree ($\rho = 0.108$, $P = 0.011$ for $\omega$; $\rho = 0.109$, $P = 4.72 \times 10^{-4}$ for $d_N$), than with out-degree ($\rho = 0.092$, $P = 0.003$ for $\omega$; $\rho = 0.092$, $P = 0.003$ for $d_N$), and they exhibit a positive, although non-significant, correlation with $H$ ($\rho = 0.048$, $P = 0.122$ for $\omega$; $\rho = 0.051$, $P = 0.104$ for $d_N$). When genes encoding kinases were also excluded from the analysis, the correlation between $H$ and rates of evolution reached significance ($\rho = 0.070$, $P = 0.046$ for $\omega$; $\rho = 0.079$, $P = 0.023$ for $d_N$). These results indicate that, although the different functions of upstream and downstream genes may partially contribute to the observed polarity in the levels of selective constraint, this factor does not completely account for the trend.

**Genes acting at the downstream part of the network exhibit higher expression levels and breadths, and encode more highly connected proteins**

Levels of selective constraint acting on a gene are affected by a number of factors, including gene expression level (measured, for instance, as the number or concentration of transcripts in the cell) and breadth (the number of tissues in which a gene is expressed) [7,50–52], codon bias [51,53], connectivity of the encoded proteins [14–16], and the number of paralogs [54,55]. Measures of these parameters were obtained from different sources (see Methods). Because the dataset generated by Cui et al. [34] focuses on signaling molecules, proteins’ total connectivities were measured as the total number of physical interactors in the whole human protein–protein interaction network [56]. All these factors significantly correlate with $\omega$ and $d_N$ in the dataset used (Table 4). Therefore, a putative polarity in the distribution of these factors across the network could potentially account for the observed polarity in the levels of selective constraint. Remarkably, expression breadth and connectivity seem to be the best correlates of the rates of evolution, with both variables exhibiting a similar degree of correlation with $\omega$ and $d_N$ (Table 4). This result contrasts with previous observations that connectivity is a relatively weak predictor of rates of evolution (e.g., refs. [5,57]), suggesting that this correlation is particularly strong for genes involved in signal transduction. Furthermore, the correlation between connectivity and rates of evolution remains significant when expression level and breadth, codon bias, and number of

### Table 4 Correlates of evolutionary rates

| Variable 1       | Variable 2       | $\rho$       | $P$         |
|------------------|------------------|--------------|-------------|
| $\omega$         | Expression level | −0.132       | 2.51×10$^{-4}$*** |
|                  | Expression breadth | −0.205       | 4.57×10$^{-1}$*** |
|                  | ENC              | 0.163        | 1.19×10$^{-4}$*** |
|                  | Connectivity     | −0.202       | 3.35×10$^{-10}$*** |
| $d_N$            | Number of paralogs | −0.116       | 1.64×10$^{-4}$*** |
|                  | Expression level | −0.141       | 6.51×10$^{-4}$*** |
|                  | Expression breadth | −0.228       | 1.76×10$^{-13}$*** |
|                  | ENC              | 0.002        | 0.947       |
|                  | Connectivity     | −0.226       | 1.68×10$^{-12}$*** |
|                  | Number of paralogs | −0.089       | 0.004**     |

***, $P<0.01$, ***, $P<0.001$.**
The lower evolutionary rates of downstream network genes are a byproduct of the distribution of protein connectivity and gene expression level and breadth across the network

I considered whether the observed polarity in the levels of selective constraint across the human signaling network was the result of the distribution of factors correlating with ω and dN. For that purpose, this polarity was evaluated after discounting the putative effects of these factors. Linear regression was used to model the dependence of ω and dN from each factor separately, and the residuals of the model for each gene (i.e., the difference between the observed and expected values) were used in the paired and the Mann–Whitney tests.

The paired tests show that genes exhibit significantly higher ω and dN values than their downstream targets even if the effect of expression level, expression breadth, codon bias, or the number of paralogs is discounted, indicating that the difference in the intensity of purifying selection acting on upstream and downstream genes is independent of these factors. However, when the effect of connectivity is discounted, the test yields no significant results (Additional file 1: Table S1).

Genes occupying extreme upstream positions in the network are significantly less constrained than those acting at the downstream part, even when the effect of expression breadth, codon bias, or the number of paralogs is removed, but the difference is not significant when expression level or connectivity are included in the analysis (Additional file 1: Table S2). Given that the relationship between evolutionary rates and their correlates is not necessarily linear in all cases, a complementary nonparametric analysis was performed to assess the differences in ω and dN values between genes occupying extreme upstream and downstream positions. Partial correlation analysis was used to evaluate the association between upstream/downstream position (encoded as a binary variable; see refs. [58,59]) and rates of evolution while controlling for each of the controlling variables. Similar results were obtained: upstream and downstream genes exhibit significantly higher ω and dN values when controlling for codon bias or the number of paralogs, but not when controlling for expression level, expression breadth, or connectivity (Additional file 1: Table S3).

Finally, partial correlation analysis was used to evaluate the association between H and ω and dN while controlling separately for each factor. Both correlations are significant when codon bias or the number of paralogs are used as controlling variables, but not when expression level, expression breadth, or connectivity are included in the analysis (Additional file 1: Table S4).

Therefore, the differences in the ω and dN values of upstream and downstream network genes disappear when the effect of expression level, expression breadth, or connectivity, are removed. This indicates that these factors account for the higher levels of selective constraint acting on the downstream genes of the human signal transduction network. In particular, when connectivities are controlled for, ω and dN values are not significantly different for upstream and downstream genes, regardless of the method used (Additional file 1: Tables...
observed in Drosophila. Most of the afore-mentioned Drosophila pathway genes have multiple orthologs in mammalian genomes known to participate in the Ras pathway (e.g., ref. [60]). For each such mammalian gene, the $\omega$ value was estimated from comparison of the human-mouse pair of orthologs (see Additional file 1: Table S5). Genes encoding the mammalian proteins SOS1, SOS2, ERK1, ERK2, and RSK1–3, which are also key components of the pathway [60], were also included in the current analysis (see Figure 2A for a diagram and description of the pathway). Contrary to previously suggested in Drosophila [37], genes acting at the upstream part of the mammalian Ras pathway (those encoding the Grb2, SOS1, SOS2, H-Ras, K-Ras, N-Ras, A-Raf, B-Raf, and C-Raf proteins) were not found to exhibit lower rates of evolution than those acting at the downstream part (the rest of genes listed in Additional file 1: Table S5). If anything, the opposite seemed to be the case (median $\omega$ for upstream and downstream genes: 0.023 and 0.018, respectively), although the differences between both groups were not significant (Mann–Whitney test, $P=0.503$). Consistently, $\omega$ values were found to exhibit a negative, although non-significant, correlation with the position of genes along the upstream/downstream axis of the pathway ($\rho=−0.195$, $P=0.470$; Figure 2B). For this analysis, pathway position was defined as the number of steps required to transduce the signal from the most upstream element (Grb2, position 0), to each of the pathway components (with proteins RSK1–3 sharing position 6, the most downstream position of the core pathway; Figure 2A; Additional file 1: Table S5).

Neither expression levels, nor expression breadths or connectivities significantly correlate with pathway position (expression level: $\rho=0.107$, $P=0.693$; expression breadth: $\rho=0.116$, $P=0.670$; connectivity: $\rho=−0.134$, $P=0.620$) or with $\omega$ values (expression level: $\rho=0.203$, $P=0.450$; expression breadth: $\rho=0.123$, $P=0.649$; connectivity: $\rho=−0.174$, $P=0.519$). It should be noted, however, that microarray-based gene expression data, such as the one used here [61,62], is subject to high levels of measurement noise. More accurate measures of expression are available for some of the proteins involved in the Ras pathway, and indeed it has been noticed that downstream proteins are present in higher concentrations in the cell (e.g., see ref. [38] and references therein), thus leaving open the possibility that the slightly higher levels of selective constraint observed in the downstream part of the pathway (Figure 2; Additional file 1: Table S5) may result from these genes being expressed at higher levels.

In any case, these results indicate that, contrary to previously suggested in Drosophila [37], genes acting at the upstream part of the mammalian Ras pathway are not subject to higher levels of selective constraint than those acting at the downstream part. If anything, the opposite might be the case, although the trend is not statistically
significant. This observation is in agreement with the general trend described here for the whole human signal transduction network.

Discussion

In summary, results presented here demonstrate that genes acting at the upstream part of the human signal transduction network clearly do not evolve under stronger selective pressures (Tables 1–3; Figure 1). This finding seems striking, as these proteins, owing to their higher hierarchical position, may be capable of regulating the activity of a higher number of downstream proteins. Therefore, mutations in the genes encoding upstream network proteins would be expected to have more pleiotropic effects, and hence these genes would be expected to evolve under stronger purifying selection. In contrast with these expectations, a number of analyses that capture different aspects of the hierarchical position of genes in the signaling network found no evidence for upstream genes being subject to higher levels of selective constraint. On the contrary, upstream network genes are less selectively constrained than downstream genes, which appears to be explained by downstream genes performing certain functions, being more highly and broadly expressed and, in particular, encoding more highly connected proteins.

Protein connectivity, gene expression breadth and, to a lesser extent, expression level, exhibit a polarity across the human signaling network, with downstream genes progressively encoding proteins that are more connected, and being expressed in a higher number of tissues and at higher levels (Tables 1–3; Figure 1). The higher levels of expression of downstream genes might facilitate the amplification of signals through their progression throughout signaling cascades, which requires higher protein abundances for the downstream elements of the pathway (e.g., ref. [38]). Additionally, the higher connectivity of downstream genes is consistent with previous observations that hubs (i.e., highly connected proteins) tend to be tightly regulated, exhibiting higher numbers of phosphorylation sites and higher mRNA decay rates [57]. The distribution of these factors across the network, combined with the fact that highly and broadly expressed genes [7,50-52], and those encoding highly connected proteins in the protein–protein interaction network [14-16], tend to evolve under stronger purifying selection, accounts for the higher levels of selective constraint observed in genes acting at the downstream part of the network. In particular, when the effect of connectivity is factored out, the differences in $\omega$ and $d_N / d_S$ values between upstream and downstream genes are no longer significant, regardless of the method used (Additional file 1: Tables S1–S4), pointing to this factor as the main responsible for the higher levels of selective constraint acting on the downstream network genes. When the effects of expression level and breadth are removed, the differences in the levels of selective constraint observed between upstream and downstream genes also vanish for some of the tests (Additional file 1: Tables S1–S4), suggesting that these factors also contribute to the tendency. Likewise, protein functions may also play a role, as the trend also vanishes for some
tests when ion transmembrane transporters (which are slow-evolving and are enriched in the downstream part of the network) are removed from the analyses. In any case, taken together these observations indicate that the upstream/downstream position of proteins in the human signal transduction network does not have a direct effect on their rates of evolution.

The global trends observed in the entire human signal transduction network (i.e., downstream genes being more selectively constrained owing to their higher connectivity, expression level and breadth) mirror those previously observed in some particular pathways. In both the Caenorhabditis insulin/TOR signaling pathway and the yeast HOG signaling pathway, downstream genes evolve under stronger levels of selective constraint, probably as a result of their higher levels of expression and codon bias (which is often used as a proxy for levels of expression), respectively [42,43]. It should be noted, however, that the general trends observed across the whole network do not necessarily imply that each particular signaling pathway must exhibit the same trend. For instance, in the Drosophila and vertebrate insulin/TOR pathways, downstream genes evolve under stronger purifying selection, even when the putative effect of several factors that correlate with levels of selective constraint is discounted [39-41,44]. Another example that does not seem to conform to the general trend described here is the Drosophila Ras pathway [37], whose upstream genes appear to be the most constrained, despite these genes being the ones expressed at lower levels [38] (however, observations in the mammalian Ras pathway do not point in the same direction; see below, Figure 2 and Additional file 1: Table S5). Therefore, despite the overall trends described here, the particular structure and function of each individual pathway may result in different distributions of the impact of evolutionary forces. For instance, pathways whose upstream genes, for some reason, present higher connectivities, and/or levels or breadths of expression, would probably be more selectively constrained in the upstream part. It should also be noted that factors other than the ones considered here may also affect the distribution of levels of selective constraint across functional pathways and networks. In particular, enzymes whose kinetic characteristics exert a high degree of influence over the entire network (and hence, on the phenotype) are expected to be more selectively constrained than those exerting little influence on the network dynamics [31,63-65]. The degree to which a particular enzyme can influence the behavior of the entire network is, however, difficult to determine from currently available data.

So far, the Drosophila Ras pathway is, to the best of the author’s knowledge, the only signal transduction pathway whose upstream genes appear to be the ones evolving under higher levels of purifying selection [37]. However, due to the limited availability of genomic data at that time, this observation was based on the analysis of a small set of genes and the comparison of two closely related species (D. melanogaster and its sister species D. simulans, which are thought to have diverged 2–6 million years ago [66,67]). Results reported here for all core components of the mammalian Ras pathway, and based on human-mouse comparisons (which probably diverged ~75 million years ago; e.g., ref. [68]), do not point in the same direction: upstream genes do not exhibit lower rates of evolution than those acting at the downstream part of the pathway. If anything, the opposite might be the case (i.e., downstream genes exhibit slightly lower rates of evolution, although no significant differences are observed between upstream and downstream genes; Figure 2; Additional file 1: Table S5). In any case, this observation lends further support to the finding that genes occupying highly hierarchical positions in the human signal transduction network are not subject to higher levels of selective constraint than those acting at the downstream part.

The observation that genes acting at the upstream part of the human signal transduction network are not subject to higher levels of selective constraint sharply contrasts with the trends generally observed in biosynthetic pathways. These pathways consist of a series of enzymes that share metabolites, with the product of a given enzyme being usually the substrate of the following enzyme (or enzymes) in the pathway. In such pathways, metabolites can often take different alternative routes, which end up in the biosynthesis of different end products. In a growing number of biosynthetic pathways, including the plant anthocyanin [28,69,70], isoprene [25], terpenoid [26], and carotenoid pathways [27,71], the Bombyx melanin pathway [30], and the mammalian dopamine pathway [29], levels of selective constraint correlate with the position of genes along the upstream/downstream pathway axis, with upstream genes being always the most constrained. Two plausible scenarios might explain such a distribution of selective constraints in biosynthetic pathways. First, as a result of their branching topology, these pathways often exhibit a hierarchical structure, with upstream genes being necessary for the biosynthesis of a higher number of end products than downstream genes; therefore, mutations in genes acting at the upstream part may have higher pleiotropic effects [28]. Second, upstream enzymes in biosynthetic pathways probably often exert a higher control over the flux of metabolites throughout the pathway [31]; thus, the overall pathway function (and, hence, the associated phenotypes) may more strongly depend on the function of these enzymes. Both models predict a higher relevance for pathway function of upstream genes, and hence a
higher strength of natural selection acting on these genes. Similar trends have been observed in transcriptional regulatory networks. First, genes occupying higher hierarchical positions in these networks are likely to be essential, reflecting their key importance for organism’s fitness [32]. Second, simulation analyses of transcriptional regulatory networks have shown that genes exerting a high degree of control over other genes, or those that are less regulated by other genes, are expected to be more strongly affected by selection [33].

Therefore, results presented here imply that the distribution of evolutionary forces across the upstream and downstream parts of the signal transduction network remarkably differs from that observed in biosynthetic and transcriptional regulatory pathways. In biosynthetic and transcriptional regulatory pathways, upstream genes evolve under stronger levels of selective constraint, whereas this is not the case for the human signal transduction network. This suggests that genes occupying higher hierarchical positions are not more relevant for the function of the signaling network, or to the organism’s fitness, than downstream genes. These different patterns of evolution may reflect fundamental differences in the function and organization of the different kinds of pathways.

As signaling networks become better understood, it becomes more apparent that signal transduction does not take place in the context of independent linear pathways in which each protein is exclusively controlled by the preceding protein in the cascade. This classical model is being progressively substituted by a much more integrative model of signal transduction, in which a considerable amount of crosstalk takes place between pathways. As a result, pathways would cooperate to integrate signals, and to make decisions about the responses in a decentralized manner, thus conferring robustness to the system [72,73]. This kind of organization may allow the network to accommodate modifications in its upstream genes. For instance, even if an upstream gene undergoes disruptive mutation, its downstream targets might still be activated by alternative routes that do not involve the disrupted element. The observation that genes occupying more hierarchical positions in the signaling network are not subject to stronger selective forces is consistent with this emerging integrative model. Although somewhat speculative, this observation would be compatible with the organization of the signal transduction network being less hierarchical than that of biosynthetic and transcriptional regulatory networks. In the future, a deeper understanding of the differences between the nature of biosynthetic, transcriptional regulatory, and signaling pathways may shed more light on their different patterns of evolution.

Conclusions
Surprisingly, genes occupying the higher hierarchical positions of the human signal transduction network are not subject to stronger levels of purifying selection, suggesting that they are not more important for the function of the network and the fitness of the organism than genes occupying the lower hierarchical positions. This observation sharply contrasts with the patterns observed in metabolic and transcriptional regulatory pathways and networks, in which upstream genes are generally the most selectively constrained. These contrasting patterns of evolution might reflect fundamental differences in the function and organization of signaling and biosynthetic and transcriptional regulatory networks. In any case, results presented here broaden our knowledge on how natural selection distributes across molecular networks.

Methods
Human signal transduction network
The human directed signaling network described in ref. [34] was used. After eliminating metabolites, proteins that could not be mapped to Ensembl IDs, proteins encoded by genes without a 1:1 ortholog in the mouse genome (see below), and undirected interactions (i.e., those for which the directionality of signal transduction is not available), the dataset contained 2436 activating or inhibitory interactions connecting 1049 human proteins.

For each human protein, a number of measures of its hierarchical position in the network were computed: in-degree and out-degree (number of upcoming and outgoing edges in the network, respectively), and the total number of proteins that lay downstream of the protein of interest (termed $H$). Proteins with high out-degrees (i.e., modulating a high number of other proteins), low in-degrees (i.e., modulated by a few or no other proteins), and/or high $H$ values (i.e., with a high number of proteins acting downstream of it) can be considered to occupy high hierarchical positions in the network. $H$ was computed as follows: first, a list of proteins containing all downstream direct targets of the protein of interest was generated; second, all direct downstream targets of these proteins were incorporated to the list. The second step was iterated until all proteins acting downstream of the protein of interest were found (i.e., until the number of elements in the list remained stable).

Impact of natural selection
All human and mouse protein and CDS sequences were obtained from the Ensembl database, release 62 [74]. For genes encoding multiple transcripts owing to alternative splicing, the variant encoding the longest protein (i.e., the “canonical transcript”) was used. For each protein in the human signaling network, the 1:1 mouse ortholog was identified using a best reciprocal BLAST approach (using BLASTP and an $E$-value cut-off of $10^{-10}$). Pairs of orthologous amino acid sequences were aligned using
ProbCons 1.12 [75], and the resulting alignments were used to guide the alignment of the corresponding CDSs. For each pair of orthologous sequences, the impact of natural selection was characterized from the nonsynonymous ($d_N$) to synonymous ($d_S$) divergence ratio ($\omega = d_N/d_S$). Estimates of $\omega$, $d_N$, and $d_S$ were obtained using the program codeml from the package PAML 4.4 [76] (under the M0 model). The action of positive selection was inferred using the M7 vs. M8 test [45]. Model M7 assumes that codons’ $\omega$ values range between 0 and 1, following a beta distribution, whereas model M8 allows for an extra class of codons with $\omega > 1$. The likelihood of both models is compared using the likelihood ratio test [77], assuming that twice the difference between the log-likelihoods of both nested models ($2\Delta\ell$) follows a $\chi^2$ distribution with 2 degrees of freedom. If model M8 significantly better fits the data, positive selection is invoked. $P$-values for this test were corrected for multiple testing using the false discovery rate approach [78].

Statistical analyses
Proteins were considered to act in the upstream part of the network if they modulate other proteins (out-degree > 0) and are not modulated by any other protein (in-degree = 0), or in the downstream part of the network if out-degree = 0 and in-degree > 0. Third, I evaluated whether the $\omega$ and $d_N$ values correlate with measures of hierarchical position in the network (in-degree, out-degree, and $H$) using the Spearman’s rank correlation coefficient ($\rho$). These methods were likewise applied to evaluate whether upstream and downstream genes differ in terms of a number of factors that are known to correlate with the levels of selective constraint: gene expression level and breadth, codon usage bias, connectivity, and number of paralogs (see below).

Furthermore, I considered whether the differences in $\omega$ and $d_N$ observed between upstream and downstream genes were attributable to the distribution of their correlates across the network. For the first two methods (paired sign and Mann–Whitney tests), linear regression was used to model the dependence of $\omega$ and $d_N$ from each factor. Then, the residuals of the model for each factor (i.e., the difference between the observed and predicted values) were used in the analyses. Additionally, partial correlation analysis was used to evaluate the correlation between rates of evolution and upstream/downstream location while controlling for each variable. For that purpose, upstream/downstream location was encoded as a binary variable (see refs. [58,59]). The Spearman’s rank correlation test, when applied to a continuous and a binary variable, is equivalent to the Mann–Whitney test. Finally, partial correlation analysis was also used to evaluate the correlation between evolutionary rates and the measures of hierarchical position (in-degree, out-degree and $H$) while controlling for each of the controlling variables.

Correlates of selective constraint levels
For each human gene, estimates for a number of factors that correlate with $\omega$ and $d_N$ were obtained from different sources:

- Gene expression level and breadth: Human gene expression data was retrieved from the BioGPS portal [61,62] (U133A/GNF1H dataset gcRMA-normalized). Probes were matched to genes through Ensembl’s BioMart [79] (for the U133A dataset) or the annotation file provided in the BioGPS portal (for the GNF1H dataset). For each probe and tissue, values were averaged across both replicates. For each probe, expression level was computed as the average across a subset of 25 nonredundant, adult noncancerous tissues (as in ref. [40]). For genes matching more than one probe, the one with the highest average across the 25 selected tissues was used. Expression breadth for each gene was calculated as the number of tissues (out of the 25 selected ones) in which the gene is expressed above the median across all tissues and genes.

- Codon bias: The effective number of codons (ENC; ref. [80]) of each gene was computed using the software CodonW 1.4.2 (http://codonw.sourceforge.net/). For each pair of human-mouse orthologs, ENC values were averaged.

- Number of paralogs: For each human gene, a list of human paralogs was obtained from Ensembl's BioMart.

- Protein connectivity: The whole human protein–protein interaction network was retrieved from the BioGRID database version 3.0.67 [56]. Only physical interactions between human proteins were used. For each protein, connectivity was computed as the number of proteins with which it physically interacts. Genes not represented in the database were assigned missing values.

- Subcellular location: For each gene, a list of the Gene Ontology [81] terms to which it is associated was obtained from Ensembl’s BioMart. Genes were considered to act in the extracellular space (105 genes), in the plasma membrane (381), cytoplasm (570) or in the nucleus (455) if associated to the terms “extracellular region”, “plasma membrane”, “cytoplasm” or “nucleus”, respectively. Among these, 34, 130, 125 and 121 act exclusively in the extracellular compartment, the plasma membrane, the cytoplasm and the nucleus, respectively.
Additional file

Additional file 1: A single PDF file containing supplementary Tables S1–S5. Table S1 lists the results of the paired tests comparing the evolutionary rates of each gene with those of its direct downstream targets while controlling for a number of correlates of rates of evolution. Table S2 lists the Mann-Whitney tests comparing the evolutionary rates of genes occupying extreme upstream and downstream positions while controlling for a number of correlates of rates of evolution. Table S3 lists the partial correlation analyses contrasting the association between upstream/downstream position of genes and their rates of evolution while controlling for a number of correlates of rates of evolution. Table S4 lists the partial correlation analyses contrasting the association between rates of evolution and measures of hierarchical position while controlling for a number of correlates of rates of evolution. Table S5 lists the $\omega$ values for genes involved in the mammalian Ras signaling pathway.

Competing interests

The author declares that he has no competing interests.

Acknowledgements

The author is indebted to Julio Rozas and James McInerney for support and excellent comments on the manuscript. He is also grateful to two anonymous referees for their comments. This work was funded in part by a Science Foundation Ireland grant to James McInerney (grant number: 09/ RFP/EGB2510), and a Juan de la Cierva fellowship from the Spanish Ministerio de Ciencia e Innovacióon to the author (JCI-2011-11089).

Publication fees were partially covered by the CSCS Open Access Publication Support Initiative through its Unit of Information Resources for Research (URSI).

Received: 25 May 2012 Accepted: 14 September 2012

Published: 28 September 2012

References

1. Li WH, Wu CI, Luo CC: A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol 1985, 2:150–174.
2. Zuckerkiand E, Pauling L: Molecules as documents of evolutionary history. J Theor Biol 1965, 8:357–366.
3. Kimura M: The neutral theory of molecular evolution. Cambridge: Cambridge University Press; 1983.
4. Koonin EV, Wolf YI: Evolutionary systems biology: links between gene evolution and function. Curr Opin Biotechnol 2006, 17:481–487.
5. Rocha EP: The quest for the universals of protein evolution. Trends Genet 2006, 22:412–416.
6. McInerney JO: The causes of protein evolutionary rate variation. Trends Ecol Evol 2006, 21:230–232.
7. Ingvarsson PK: Gene expression and protein length influence codon usage and rates of sequence evolution in Populus tremula. Mol Biol Evol 2007, 24:835–844.
8. Butcher EC, Berg EL, Kunkel EJ: Systems biology in drug discovery. Nat Biotechnol 2004, 22:1253–1259.
9. Lee JW, Kim TY, Jang YS, Choi S, Lee SY: Systems metabolic engineering for chemicals and materials. Trends Biotechnol 2011, 29:370–378.
10. Koncaroñs T, Szalay MS, Böde C, Kovács IA, Csermely P: How to design multi-target drugs: target search options in cellular networks. Expert Opinion on Drug Discovery 2007, 2:799–808.
11. Cork JM, Purugganan MD: The evolution of molecular genetic pathways and networks. Bioessays 2004, 26:479–484.
12. Zera AJ: Microevolution of intermediary metabolism: evolutionary genetics meets metabolic biochemistry. J Exp Biol 2011, 214:179–190.
13. Eanes WF: Molecular population genetics and selection in the glycolytic pathway. J Evol Biol 2011, 214:165–171.
14. Fraser HB, Hirth AE, Steinitz LM, Scharfe C, Feldman MW: Evolutionary rate in the protein interaction network. Science 2002, 296:750–752.
15. Hahn MW, Kern AD: Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. Mol Biol Evol 2005, 22:803–806.
16. Lemos B, Bettencourt BR, Meiklejohn CD, Hartl DL: Evolution of proteins and gene expression levels are coupled in Drosophila and are independently associated with mRNA abundance, protein length, and number of protein-protein interactions. Mol Biol Evol 2005, 22:1345–1354.
17. Vitkup D, Kharchenko P, Wagner A: Influence of metabolic network structure and function on enzyme evolution. Genome Biol 2006, 7: R39.
18. Greenberg AJ, Stockwell SR, Clark AG: Evolutionary constraint and adaptation in the metabolic network of Drosophila. Mol Biol Evol 2008, 25:2537–2546.
19. Hudson OM, Cantar GC: Expression level, cellular compartment and metabolic network position all influence the average selective constraint on mammalian enzymes. BMC Evol Biol 2011, 11:89.
20. Cui Q, Purisma EO, Wang E: Protein evolution on a human signaling network. BMC Syst Biol 2009, 3:21.
21. Lovell SC, Robertson DL: An integrated view of molecular coevolution in protein-protein interactions. Mol Biol Evol 2010, 27:2567–2575.
22. Cardenas FM, Faere MA: Why should we care about molecular coevolution? Evol Bioinform Online 2008, 4:29–38.
23. Alvarez-Ponce D, McInerney JO: The human genome retains relics of its prokaryotic ancestry: human genes of archaeabacterial and eubacterial origin exhibit remarkable differences. Genome Biol 2011, 3:782–790.
24. Doherty A, Alvarez-Ponce D, McInerney JO: Increased genome sampling reveals a dynamic relationship between gene duplicability and the structure of the primate protein-protein interaction network. Mol Biol Evol, in press.
25. Sharkey TD, Yeh S, Wiberley AE, Falbel TG, Gong D, Fernandez DE: Evolution of the isoprene biosynthetic pathway in kudzu. Plant Physiol 2005, 137:700–712.
26. Ramsay H, Rieseberg LH, Ritland K: The correlation of evolutionary rate with pathway position in plant terpenoid biosynthesis. Mol Biol Evol 2009, 26:1045–1053.
27. Livingstone K, Anderson S: Patterns of variation in the evolution of carotenoid biosynthetic pathway enzymes of higher plants. J Hered 2009, 100:754–761.
28. Rausher MD, Miller RE, Tiffin P: Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. Mol Biol Evol 1999, 16:266–274.
29. Ma X, Wang Z, Zhang X: Evolution of dopamine-related systems: biosynthesis, degradation and receptors. J Mol Evol 2010, 71:374–384.
30. Yu HS, Shen YH, Yuan GX, Hu YG, Xu HE, Xiang ZH, Zhang Z: Evidence of selection at melanin synthesis pathway loci during silkworm domestication. Mol Biol Evol 2011, 28:1785–1799.
31. Wright KM, Rausher MD: The evolution of control and distribution of adaptive mutations in a metabolic pathway. Genetics 2010, 184:483–502.
32. Bhadradw N, Kim PM, Gerstein MB: Rewiring of transcriptional regulatory networks: hierarchy, rather than connectivity, better reflects the importance of regulators. Sci Signal 2010, 3ra79.
33. Rhone B, Brandenburg JT, Austerlitz F: Impact of selection on genes involved in regulatory network: a modelling study. J Mol Evol 2011, 24:2087–2098.
34. Cui Q, Ma Y, Jaramillo M, Bari H, Awan A, Yang S, Zhang S, Liu L, Lu M, O'Connor-McCourt M, et al: A map of human cancer signaling. Mol Syst Biol 2007, 3:152.
35. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M: KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 1999, 27:29–34.
36. Ma'ayan A, Jenkins SL, Hasseldine A, Graze E, Dubin-Thaler B, Euskirchen GM, Weng G, Ram PT, Rice DJ, et al: Formation of regulatory patterns during signal propagation in a Mammalian cellular network. Science 2005, 309:1078–1083.
37. Riley RM, Jin W, Gibson G: Contrastting selection pressures on components of the Ras-mediated signal transduction pathway in Drosophila. Mol Syst Biol 2003, 12:1315–1323.
38. Ferrell JE Jr: Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switch-like outputs. Trends Biochem Sci 1996, 21:460–466.
39. Alvarez-Ponce D, Agudelo M, Rozas J: Network-level molecular evolutionary analysis of the insulin/TOR signal transduction pathway across 12 Drosophila genomes. Genome Res 2009, 19:234–242.
40. Alvarez-Ponce D, Aguade M, Rozas J: Comparative genomics of the vertebrate insulin/TOR signal transduction pathway: a network-level analysis of selective pressures. Genome Biol 2011, 12:R71.

41. Alvarez-Ponce D, Guinou-Rico S, Orrego DJ, Segarra C, Rozas J, Aguade M: Molecular population genetics of the insulin/TOR signal transduction pathway: a network-level analysis in Drosophila melanogaster. Mol Biol Evol 2012, 29:123–132.

42. Jovein R, Phillips EC: Expression level drives the pattern of selective constraints along the insulin/TOR signal transduction pathway in Caenorhabditis. Genome Biol 2011, 12:715–722.

43. Wu X, Chi X, Wang P, Zheng D, Ding R, Li Y: The evolutionary rate variation among genes of HOG-signaling pathway in yeast genomes. Biol Direct 2010, 5:46.

44. Luisi P, Alvarez-Ponce D, Dall’olio GM, Sikora M, Bertranpetit J, Laayouni H: The functional genomic distribution of protein divergence in two animal phyla: coevolution, genomic conflict, and contrast. Genome Res 2004, 14:802–811.

45. Al-Shahour F, Minguez P, Tartaglia J, Medina I, Alloza E, Montaner D, Dopazo J: FatigO +: a functional profiling tool for genomic data. Integration of models of sequence evolution in phylogenetics. Mol Biol Evol 2007, 24:1586–1591.

46. Tamura K, Subramanian S, Kumar S: Temporal patterns of fruit fly (Drosophila) evolution revealed by mutation clocks. Mol Biol Evol 2004, 21:36–44.

47. Russo CA, Takezaki N, Nei M: Molecular phylogeny and divergence times of drosophilid species. Mol Biol Evol 1995, 12:391–404.

48. Ponting CP: The functional repertoires of metazoan genomes. Nat Rev Genet 2008, 9:689–698.

49. Lyu Y, Rauscher MD: Evolutionary rate variation in anthocyanin pathway genes. Mol Biol Evol 2003, 20:1844–1853.

50. Rausher MD, Lyu Y, Meyer K: Variation in constraint versus positive selection as an explanation for evolutionary rate variation among anthocyanin genes. J Mol Evol 2008, 67:147–159.

51. Clatot J, Pellet D, Soulhuet-Feslon V, Briard M, Geoffroy E: Differential selection on carotenoid biosynthesis genes as a function of gene position in the metabolic pathway: a study on the carrot and dicots. PLoS One 2012, 7:e38724.

52. Levy ED, Landry CR, Michnick SW: Cell signaling: Signaling through cooperation. Science 2010, 328:983–984.

53. Breitkreutz A, Choi H, Sharom JR, Boucher L, Neduva V, Larsen B, Lin ZY, Breitkreutz BJ, Stark C, Liu G, et al: A global protein kinase and phosphatase interaction network in yeast. Science 2010, 328:1043–1046.

54. Flicek P, Amode MR, Barrell D, Besil K, Brent S, Chen Y, Clapham P, Coates G, Fairley S, Fitzgerald S, et al: Ensembl 2011. Nucleic Acids Res 2011, 39:D800–806.

55. Do CB, Mahabhashyam MS, Buduco M, Bhatia J: ProBcon: Probabilistic consistency-based multiple sequence alignment. Genome Res 2005, 15:330–340.

56. Yang Z: PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 2007, 24:1586–1591.

57. Whelan S, Goldman N: Distributions of statistics used for the comparison of models of sequence evolution in phylogenetics. Mol Biol Evol 1999, 16:292–299.

58. Benjamin Y, Hochberg Y: Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc B 1995, 57:289–300.

59. Kinsella RJ, Kahari A, Haider S, Zamora J, Proctor G, Spudich G, Almeida-King J, Staines D, Derwent P, Kerhornou A, et al: BioGPS: an extensible and customizable resource which is freely available for redistribution. Nucleic Acids Res 2011, 39:10155–10161.

50. Duret L, Mouchiroud D: Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. Mol Biol Evol 2000, 17:688–74.

51. Liao BY, Weng MP, Zhang J: Highly expressed genes in yeast evolve slowly. Genetics 2001, 158:277–381.

52. Subramanian S, Kumar S: Gene expression intensity shapes evolutionary rates of the proteins encoded by the vertebrate genome. Genetics 2004, 168:373–381.

53. Sharp PM: Determinants of DNA sequence divergence between Escherichia coli and Salmonella typhimurium: codon usage, map position, and concerted evolution. J Mol Biol 1991, 232:23–33.

54. Lynch M, Conery JS: The evolutionary fate and consequences of duplicate genes. Science 2000, 290:1151–1155.

55. Jordan IK, Wolf YI, Koonin EV: Duplicated genes evolve slower than singletons despite the initial rate increase. BMC Biol Evol 2004, 4:22.

56. Breitkreutz BJ, Stark C, Reguly T, Boucher L, Breitkreutz A, Livstone M, Oughtred R, Lackner DH, Bahler J, Wood V, et al: The BioGRID Interaction Database: 2008 update. Nucleic Acids Res 2008, 36:D637–640.

57. Batada NN, Hurst LD, Tyers M: Evolutionary and physiological importance of hub proteins. PLoS Comput Biol 2006, 2:e68.

58. Hardy MA: Regression with dummy variables. London: Sage Publications; 1993.

59. Cohens J: Applied multiple regression/correlation analysis for the behavioral science. 3rd edition. Mahwah, New Jersey: Lawrence Erlbaum; 2003.

60. McCay MM, Morrison DK: Integrating signals from RTKs to ERK/MAPK. Oncogene 2007, 26:3113–3121.

61. Su AI, Wittholt B, Batalov S, Lapp H, Chang KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, et al: A gene atlas of the mouse and human protein-encoding transcripts. Proc Natl Acad Sci U S A 2004, 101:6062–6067.

62. Wu C, Orozco C, Boyer J, Leglise M, Goodale J, Batalov S, Hodge CL, Haase J, James J, Huss JW 3rd, Su AI: BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. Genome Biol 2012, 2:123–132.

63. Hartl DL, Doherty SE, Dean AM: Limits of adaptation: the evolution of selective neutrality. Genetics 1985, 111:655–674.

64. Eanes WF: Analysis of selection on enzyme polymorphisms. Rev Ecol Syst 1999, 30:301–326.

65. Watt WB, Dean AM: Molecular-functional studies of adaptive genetic variation in prokaryotes and eukaryotes. Annu Rev Genet 2000, 34:593–622.