Network-Level and Population Genetics Analysis of the Insulin/TOR Signal Transduction Pathway Across Human Populations

Pierre Luisi, David Alvarez-Ponce, Giovanni Marco Dall'Olio, Martin Sikora, Jaume Bertranpetit, and Hafid Laayouni

1Institute of Evolutionary Biology (UPF-CSIC) CEXS-UPF-PRBB, Barcelona, Catalonia, Spain
2Departament de Genètica, Universitat de Barcelona, Barcelona, Spain
3Department of Biology, National University of Ireland Maynooth, Maynooth, County Kildare, Ireland
4Department of Genetics, Stanford University School of Medicine

Abstract

Genes and proteins rarely act in isolation, but they rather operate as components of complex networks of interacting molecules. Therefore, for understanding their evolution, it may be helpful to take into account the interaction networks in which they participate. It has been shown that selective constraints acting on genes depend on the position that they occupy in the network. Less understood is how the impact of local adaptation at the intraspecific level is affected by the network structure. Here, we analyzed the patterns of molecular evolution of 67 genes involved in the insulin/target of rapamycin (TOR) signal transduction pathway. This well-characterized pathway plays a key role in fundamental processes such as energetic metabolism, growth, reproduction, and aging and is involved in metabolic disorders such as obesity, insulin resistance, and diabetes. For that purpose, we combined genotype data from worldwide human populations with current knowledge of the structure and function of the pathway. We identified the footprint of recent positive selection in nine of the studied genomic regions. Most of the adaptation signals were observed among Middle East and North African, European, and Central South Asian populations. We found that positive selection preferentially targets the most central elements in the pathway, in contrast to previous observations in the whole human interactome. This observation indicates that the impact of positive selection on genes involved in the insulin/TOR pathway is affected by the pathway structure.

Key words: human populations, positive selection, insulin/TOR signal transduction pathway, network evolution.

Introduction

During their diaspora, humans have been exposed to a wide range of different environments, to which they have had to adapt. To study the process of adaptation, we can interrogate genomic data to determine which genes evolved under positive selection. Since positive selection leads to the adaptation of a specific phenotype to a given environment, such knowledge may be helpful to understand the complex phenotype–genotype relationship, which remains largely unknown. Whereas comparison of the genomes of different species can provide insight into the evolutionary forces that acted at a large timescale, using data from individuals of the same species enables the detection of the genomic footprints of more recent adaptive events.

Genes in a genome are affected by disparate evolutionary forces, and one of the goals in Evolutionary Biology is to determine the factors underlying this variability. Genes and proteins rarely act in isolation but rather operate as components of complex networks of interacting molecules. Therefore, for understanding their evolution, it may be helpful to take into account the interaction networks in which they participate (Loewe 2009; Yamada and Bork 2009; Barton et al. 2010; Wolf et al. 2010). The combination of comparative genomics and network data has shown that the patterns of molecular evolution of genes are related to the structure of the networks in which they participate, with a number of measures of network position correlating with levels of selective constraint. First, genes occupying more central positions in a protein–protein interaction network are more selectively constrained than those acting at the periphery (Fraser et al. 2002; Hahn and Kern 2005; Lemos et al. 2005). Second, genes encoding interacting proteins are similarly affected by purifying selection (Fraser et al. 2002; Lemos et al. 2005), which can be attributed to molecular coevolution or to interacting genes being affected by similar selective pressures (Codóner and Fares 2008; Fares et al., 2011). And third, in a number of pathways, including the insulin/TOR pathway (Alvarez-Ponce et al. 2009; Alvarez-Ponce, Aguade et al. 2011; Alvarez-Ponce, Guirao-Rico et al. 2011), a polarity in the distribution of levels of selective constraint levels has been observed along the upstream/downstream pathway axis (Rausher et al. 1999; Riley et al. 2003; Sharkey et al. 2005; Livingstone and Anderson 2009; Ramsay et al. 2009).
Less understood is how the patterns of positive selection are affected by the position of genes in the network, although some tendencies have been observed: positive selection seems to preferentially target genes encoding the most peripheral proteins in the human protein–protein interaction network (Kim et al. 2007), and in Drosophila, positive selection tends to target genes acting at branch points of the pathways involved in glucose metabolism (Flowers et al. 2007). More work is required to understand the relationship between the structure of functional networks and the action of positive selection.

Molecular pathways can be classified as metabolic, signal transduction, and transcriptional regulatory pathways. Signaling pathways allow the integration of cell function with extracellular stimuli. The insulin/TOR signal transduction pathway plays a key role in controlling energetic metabolism, growth, reproduction, and aging in response to nutritional conditions (Oldham and Hafen 2003; LeRoith et al. 2004). The structure and function of this pathway have been well characterized in a number of organisms, revealing a high degree of conservation across metazoans (Oldham and Hafen 2003; Kirkness et al. 2010). Insulin is secreted in response to high levels of blood sugar. Upon binding to insulin, the insulin receptor (Inr) undergoes autophosphorylation, thus becoming able to recruit a complex of proteins to the inner cell membrane, including the phosphatidylinositol 3-kinase (PI3K). PI3K catalyzes the biosynthesis of phosphatidylinositol 3,4,5-trisphosphate (PIP3), a membrane lipid that serves as anchor for pleckstrin homology domain-containing proteins, including PDK1, PKB, and PKC. Once in the membrane, these proteins become able to trigger the activation of downstream proteins, including transcription factors and ribosomal proteins, which in turn mediate the physiological effects of insulin.

The prevalence of obesity and closely related diseases such as insulin resistance and type 2 diabetes is dramatically different among human populations (King et al. 1998; Reiner et al. 2007; van Vliet-Ostaptchouk et al. 2009). The differences observed among individuals of different ethnic origin in the same country (Diamond 2003; Wang and Beydoun 2007; Anderson and Whitaker 2009; Diamond 2011) underline the fact that these metabolic disorders are not only caused by environmental exposures but also by genetic factors. The insulin/TOR pathway plays a key role in these diseases; indeed, dysfunction of its components often leads to insulin resistance, diabetes, obesity, or cancer (Oldham and Hafen 2003; LeRoith et al. 2004; Taguchi and White 2008). Thus, applying population genetics analysis to describe how natural selection acted in different populations on the genes involved in this pathway may provide key insight into the etiology of these diseases.

Here, we used human population genetics data to study the impact of recent positive selection (within the last ~5,000–45,000 years) on the molecular evolution of 67 genes involved in the human insulin/TOR signal transduction pathway. We identified signals of potential recent positive selection in nine of the studied genes. Most of these signals were observed in Middle East and North African, European, and Central South Asian populations. We found that the impact of positive selection on the insulin/TOR pathway genes depends on the position that their encoded products occupy in the interaction network. Contrary to previously observed in the human protein–protein interaction network (Kim et al. 2007), positive selection preferentially targets genes encoding the most central proteins in the pathway. This result indicates that the structure of the pathway influences the patterns of molecular evolution of its components.

Materials and Methods
Genomic Regions Analyzed
We obtained a list of 69 human genes from a previous study of the vertebrate insulin/TOR pathway (Alvarez-Ponce, Aguade et al. 2011). Most of these genes are known to be involved in the pathway according to the literature or are close paralogs that cluster within these genes by phylogenetic trees. We excluded from the analysis genes that, despite having paralogs that are involved in the pathway, are known not to play a role in insulin/TOR signaling: INSRR, encoding the insulin-related receptor, which is not activated by insulin or IGFs (Zhang and Roth 1992); and PIK3CG, encoding the protein p110γ, which is activated by G protein–coupled receptors and not by tyrosine kinases (Stephens et al. 1997; Leopoldt et al. 1998). Additionally, we included in the present analysis the genes encoding insulin (INS) and the insulin-like growth factors I (IGF1) and II (IGF2). Some of the methods used for detecting signals of positive selection (see below) have been devised for autosomal regions or provide results that cannot be compared between the autosomal regions and the ones linked to the X chromosome. Therefore, genes IRS4, FOXO4, and MYCL2, which are linked to chromosome X, were excluded from the analysis. The final list therefore consisted of 67 genes, which belong to 24 groups of paralogs (supplementary table S1, Supplementary Material online).

For each gene, we analyzed the genomic region including the gene plus 200 flanking kilobases (100 Kb upstream and 100 Kb downstream of the transcript). Gene coordinates were obtained from release 36 of the human genome at NCBI. For genes with multiple splicing variants, we considered the transcript spanning the longest chromosome region. Two pairs of overlapping regions (INS and IGF2; PDK1 and AC141586.2) were merged, and therefore, the analyses were conducted on a total of 65 genomic regions (supplementary table S1, Supplementary Material online), of which 58 were used in network-level analyses (see below).

Population Genomic Data
For each of the 65 studied genomic regions, we obtained single nucleotide polymorphism (SNP) data from the Human Genome Diversity Panel–Centre d’Etudes du Polymorphisme Humain (HGDP-CEPH) (Cann et al. 2002). We restricted our analysis to a subset composed of 940 genotyped individuals from 39 populations as described...
in Gonzalez-Neira et al. (2006). This subpanel does not contain any pair of individuals with a first or second degree relationship (Rosenberg 2006). These populations were grouped into seven geographic regions as defined by Li et al. (2008): Sub-Saharan Africa, Middle East and North Africa, Europe, Central South Asia, East Asia, America, and Oceania. These seven groups differ in both the number of populations (ranging from two to nine) and the number of individuals (from 27 to 219; supplementary table S2, Supplementary Material online). For each individual, data for ~650,000 SNPs distributed along the whole genome is available (Jakobsson et al. 2008; Li et al. 2008).

We removed all SNPs with a minor allele frequency (MAF) lower than 5% either considering individuals within each geographic area or considering all individuals from the seven geographic regions together, depending on whether the methods were applied within a specific geographic area or to compare different regions, respectively. After filtering at a global level, our data set consists of 586,210 SNPs (the number of SNPs used for separate analysis of each geographic area is provided in supplementary table S3, Supplementary Material online).

Detection of Positive Selection
In order to detect signals of positive selection in the studied genomic regions, three different methods were used. The integrated haplotype score (iHS) is based on the extended haplotype homozygosity measure, which aims at detecting the footprint of positive selection from the local haplotype structure. The two other methods (\(F_{ST}\) and \(\Delta\)DAF indexes) are based on genetic differentiation between geographic areas.

A selective sweep causes the selected allele to spread rapidly through the population, leading to an increased linkage disequilibrium compared with the neutral expectation. In each geographic region, we computed a raw iHS for each SNP following the method proposed by Voight et al. (2006). Standardized iHS were obtained by grouping SNPs into bins of frequency 0.05, subtracting the mean, and dividing by the standard deviation for all SNPs in the same bin. Extreme positive or negative values indicate high extended haplotype homozygosity of haplotypes carrying the ancestral or derived allele, respectively. Hence, we consider both extreme positive or negative iHS as potential signatures of positive selection, using \(|\text{iHSs}|\) as in Voight et al. (2006). Genetic distances were obtained from the genetic map provided by the HapMap Consortium release 22 (International HapMap Consortium 2007). Genotype phases were inferred using the fastPhase program (Scheet and Stephens 2006).

The \(F_{ST}\) index (Weir and Hill 2002) summarizes genetic differentiation among populations by estimating the proportion of interindividual variance that can be explained by the intergroup variance. For each SNP, we computed the \(F_{ST}\) index between each geographic region and the remaining ones using the PopGen module from BioPerl (Stajich et al. 2002). \(\Delta\)DAF (Grossman et al. 2010) describes the difference in derived allele frequencies (DAF) between two sets of individuals. This score was computed as \(D_S - D_{NS}\), where \(D_S\) and \(D_{NS}\) are the DAF in the studied geographic region and in the remaining ones, respectively. \(\Delta\)DAF scores range from \(-1\) to \(1\), with negative and positive values indicating whether the derived allele is less or more frequent in the given geographic area, respectively. Large \(F_{ST}\) and \(\Delta\)DAF values can be the result of the action of positive selection on the studied SNP or in a linked one in a specific geographic area (Lewontin and Krakauer 1973; Grossman et al. 2010). Computation of the iHS and \(\Delta\)DAF score requires ancestral allele information. Ancestral states inferred from comparison with orthologous sequences in the chimpanzee and rhesus macaque genomes were obtained from the UCSC Genome Bioinformatics Site (http://genome.ucsc.edu/; table ‘snp128OrthoPanTro2RheMac2’; Karolchik et al. 2008).

Extreme values for the used statistics can also be the result of nonselective events such as demographic changes and genetic drift. As these events act on the genome randomly, in contrast to positive selection, which targets specific genes, we adopted an outlier approach to infer the action of positive selection (Kelley et al. 2006). Once scores from the three methods were calculated for all SNPs in the genome, we computed empirical \(P\) values for the SNPs in the studied genomic regions as in Kelley et al. (2006). Under this approach, the \(P\) value corresponds to the probability that a value randomly drawn from the whole genome is more extreme than the value observed at the SNP of interest. If the locus of interest shows an outlier value (i.e., less than 5% of the SNPs from the background present a greater value than the SNP of interest), positive selection is invoked. Because differentiation indexes (Beaumont and Nichols 1996; Barreiro et al. 2008; Gardner et al. 2008; Jost 2008) strongly correlate with heterozygosity, we compared these indexes with the ones observed at loci with similar MAF values. For that purpose, we divided the background SNP set into bins of 10,000 SNPs of similar MAF values, and \(P\) values for each SNP were computed as the proportion of SNPs in the same bin with a greater score.

Having calculated \(P\) values separately for each SNP of interest, we integrated results for all SNPs within each of the studied genomic regions. For that purpose, we performed the Fisher’s combination test (Zaykin et al. 2008) for each genomic region, geographic area, and method. The statistic is computed as \(Z_F = -2 \sum_{i=1}^{K} \log P_i\), where \(K\) is the number of SNPs in the studied genomic region and \(P_i\) is the empirical \(P\) value for the SNP \(i\). This statistic follows a \(\chi^2\) distribution with \(2K\) degrees of freedom. This method shows a good performance for combining information from SNPs even if they are not completely independent (Peng et al. 2009). Moreover, the HGDP-CEPH data set was generated using Illumina arrays that mostly contain tag SNPs, which capture most of the information from all SNPs in the region by linkage disequilibrium, and hence present a low level of linkage disequilibrium among them. After obtaining the \(P\) values from this Fisher’s combination test, we evaluated their
significance taking into account the whole-genome context. For that purpose, we used a genomic region-level background containing the 65 genomic regions encoding the insulin/TOR pathway plus a set of 6,329 nonoverlapping regions centered on autosomal genes distributed across the genome. The complete background gene set obtained thus includes 6,394 genomic regions and 386,401 SNPs. To avoid overlap among background genomic regions, they were selected in each autosomal chromosome as follows: beginning from the gene at the lowest physical position, genes located at a minimum distance of 200 Kb from the previous one were included until reaching the end of the chromosome. From this set of genes, the genomic regions including the genes plus 200 flanking kilobases were considered for the analysis, as genomic regions were constructed for insulin/TOR pathway genes. For each of these background genomic regions, we also computed a P value using the Fisher’s combination test. P values were corrected for multiple testing using the false discovery rate approach (Benjamini and Hochberg 1995). Genomic regions were considered to have evolved under positive selection if they had a corrected P value lower than a given threshold in any of the studied regions for iHS and, at least, one of the genetic differentiation-based methods used (i.e., either $F_{ST}$ or $\Delta DAF$). Since we are considering three methods and seven geographic regions, we used a highly conservative threshold of $2.381 \times 10^{-3}$ (i.e., 0.050/$3 \times 7$).

**Network-Level Analysis**

We evaluated whether the structure of the insulin/TOR pathway has an effect on the molecular evolution of its components. For that purpose, we considered whether 1) there is a polarity in the impact of positive selection along the upstream/downstream axis of the pathway; 2) the incidence of positive selection is different for central and peripheral genes in the pathway; and 3) genes encoding physically interacting proteins evolve under similar selective pressures.

For these analyses, we used the graph constructed by Alvarez-Ponce, Aguade et al. (2011), which we expanded by adding an additional node representing insulin and the insulin-like growth factors I and II (IGF1 and IGF2), and an arch representing the activation of the insulin and IGF1 receptors (InR and IGF1R) by these hormones. As in Alvarez-Ponce, Aguade et al. (2011), we eliminated from network-level analyses: 1) genes EIF4E2 and EIF4E3, as their encoded products likely act as negative regulators of insulin signaling (Joshi et al. 2004; Hernandez et al. 2005); 2) PTEN, as its encoded product does not interact with any other protein in our data set (for a review, see Vinciguerra and Foti 2006); and 3) genes CYTH1–4, whose encoded products occupy an unclear position in the pathway (Fuss et al. 2006; Hafner et al. 2006). Hence, network-level analyses were conducted on a total of 60 genes located at 58 genomic regions; and the final graph consists of 22 nodes (representing groups of paralogous genes) connected by 40 arcs (interactions), of which 33 are physical (fig. 1).

We first tested for a polarity in the action of positive selection along the pathway by comparing pathway positions of genes evolving under positive selection with positions of genes without signals of adaptive evolution. For each gene, pathway position, defined as the number of steps required for the transduction of the signal from InR and IGF1R (position 0) to each of the components of the pathway (positions 1–10), was computed as in Alvarez-Ponce, Aguade et al. (2011). Genes INS, IGF1, and IGF2 were assigned the position −1, as their encoded products act immediately upstream of these receptors (see fig. 1).

Second, we evaluated whether genes evolving under positive selection differ from those without signals of adaptive evolution in terms of a number of metrics of centrality within the network: connectivity, betweenness and closeness (table 1). For each node, connectivity was computed as the number of nodes to which it is connected, betweenness was calculated as the number of shortest paths passing through it, and closeness as the inverse of the average length of the shortest paths to all the other nodes in the network. These metrics were computed using the Network Analysis plug-in for Cytoscape (Smoot et al. 2011), taking into account physical interactions among the studied proteins only. Additionally, these centrality measures were also computed taking into account the entire human interactome constructed by Bossi and Lehner (2009) and removing self-interactions (data not shown).

Finally, we investigated whether genes evolving under positive selection encode proteins that tend to interact to each other. To that end, we performed a Monte Carlo test using as statistic the number of interactions involving two paralogous groups (nodes) with at least one gene evolving under positive selection (X). The statistical significance of X was evaluated by comparing the observed values with those calculated on a set of 10,000 randomizations of the network, each with the same nodes and the same number of interactions as the original one. We used two different randomization techniques. In method 1, each interaction was assigned by randomly choosing two different nodes (method 1). Additionally, we used a switching algorithm, which preserves the connectivity of each particular node (method 2). Each random network was generated from the original one by repeatedly choosing two interactions at random (e.g., A–B and C–D) and exchanging the edges (yielding A–D and C–B or A–C and B–D). In each random network, this process was iterated $100 \times m$ times, where $m$ is the number of edges. $P$ values were computed as the proportion of randomized networks with an X value equal to or higher than the observed one. We also performed these analyses using as statistic the number of interactions involving the encoded products of two particular genes, rather than paralogous groups, evolving under positive selection (X’). For these analyses, we assumed that, if two groups of paralogs interact, all proteins encoded by genes in one group interact with all proteins encoded by genes in the other group.
Action of Positive Selection on Paralogous Genes

To contrast whether genes belonging to the same paralogous group are similarly targeted by positive selection, we used a Monte Carlo method using as statistic the number of pairs of paralogs in which both exhibit signals of positive selection (Y). We tested whether the observed Y value is higher than expected at random from 10,000 random sets of paralogous groups. Each randomization had the same 65 genes and the same number of pairs of paralogs (112 pairs). Each pair of paralogs was assigned by randomly choosing two different genes.

Results

High Incidence of Positive Selection in the Insulin/TOR Pathway

We used population genetics data from seven human worldwide geographic areas to study the patterns of molecular evolution of the genomic regions containing 67 autosomal genes involved in the insulin/TOR signal transduction pathway. In total, we analyzed 4,442 SNPs spanning 19.3 Mb of the human genome (supplementary table S1, Supplementary Material online). After removing SNPs with MAF lower than 5% in the entire data set, 4,005 SNPs were used in the study, providing an average density of approximately 1 SNP/5 Kb.

We combined three complementary methods (iHS, FST, and ΔDAF) to infer the action of positive selection in these genomic regions. Supplementary fig. S1–S3, Supplementary Material online show, for each continental group, the iHS, ST, and ΔDAF scores, respectively, for the SNPs located in large genomic regions centered on the genes of interest. These plots allow comparing observations for SNPs within the gene and nearby with SNPs in the surrounding region, thus, they provide a better visualization of potential signals of positive selection. Using a stringent criterion (see Materials and Methods), we identified footprints of potential positive selection in nine genes belonging to five different groups of paralogs (table 2, supplementary tables S4–S6, Supplementary Material online). For six of these genes, the signature of positive selection was observed in a single geographic area, whereas for the three other genes positive selection has acted in two distinct geographic regions, always from Middle East and North Africa, Europe, or Central South Asia regions.

First, we observed the footprint of positive selection for three genes encoding docking proteins (DOK1, DOK2, and DOK5), belonging to the IRS paralogous group. DOK1 fulfills our criterion in Middle East and North Africa.
and in Europe and presents also a significantly high genetic differentiation between East Asia and the rest of the regions according to both the $F_{ST}$ and the $\Delta DAF$ scores. $DOK2$ exhibits the signature of positive selection in Central South Asia, as well as significantly high $F_{ST}$ and $\Delta DAF$ scores between Sub-Saharan Africa and the other regions. We also observed significant $iHS$ at $DOK2$ in Middle East and North Africa and in Europe. $DOK5$ meets the criterion in Middle East and North Africa and in Central South Asia and shows a significant degree of genetic differentiation in Sub-Saharan Africa (for both the $F_{ST}$ and the $\Delta DAF$ scores), Europe ($\Delta DAF$), and East Asia ($F_{ST}$). From the PKB group, $AKT1$ and $AKT3$ satisfy our criterion in Europe and Middle East and in North Africa, respectively. Both genes also exhibit a significantly high $F_{ST}$ score ($AKT1$ for East Asia and $AKT3$ for Sub-Saharan Africa). Moreover, we detected signatures of positive selection for two genes that belong to the PKC paralogous group: $PRKCI$ (in Oceania) and $PRKCH$ (in Middle East and North Africa and in Europe). For the latter gene, we also identified an important degree of genetic differentiation between Sub-Saharan Africa and the other regions and a significantly high $iHS$ in Central South Asia and in East Asia. $RPS6KB2$, from the S6K group, exhibits signatures of adaptive evolution in Central South Asia and a significantly high relevant genetic differentiation of the Sub-Saharan Africa area, as well as significant $iHS$ in Middle East and North Africa and in East Asia. Finally, we observed the signature of positive selection in East Asia for $EIF4E$.

It is interesting to note that for the seven genes for which we observed signals of positive selection in Middle East and North Africa, Europe, or Central South Asia ($DOK1$, $DOK2$, $DOK5$, $AKT1$, $AKT3$, $PRKCH$, and $RPS6KB2$), the selected haplotypes seem to have spread through all three areas. Indeed, in most cases, these genes meet our criterion in two geographic regions or at least exhibit a statistically significant $iHS$ in two or three areas. Moreover, distributions along the genomic regions of the $iHS$ at the SNP level are very similar in these geographic areas (e.g., see $DOK1$ and $DOK2$; supplementary fig. S1, Supplementary Material online). We carried out a complementary analysis using the software Sweep 1.1 (Sabeti et al. 2002), which indicates that the haplotypes for which we detect a selective sweep are shared between the three areas (see supplementary results, Supplementary Material online), which is consistent with a genome-wide scan performed by Pickrell et al. (2009).

We considered whether there is an enrichment in genes with signals of positive selection among the insulin/TOR signal transduction pathway genes as compared with the set of 6,394 genomic regions that compose our genomic background. To that end, we performed a hypergeometric test in each geographic area separately (table 3). We observed that there is an overrepresentation of genes with the signature of positive selection in the insulin/TOR pathway genes in Middle East and North Africa (four genes under positive selection; $P = 0.0033$), Europe (three genes; $P = 0.0167$), and Central South Asia (three genes; $P = 0.0184$).

### Influence of Network Structure on Positive Selection

Having identified the genes involved in the insulin/TOR signal transduction pathway that likely evolved under positive selection (table 2), we investigated the relationship between the patterns of evolution of the studied genes and the position that their encoded products occupy in the network. For that purpose, we first evaluated whether the nine genomic regions showing footprints of positive selection differed from the 49 remaining ones in terms of centrality within the network and position along the upstream/downstream pathway axis. The position of genes along the pathway (defined as the number of steps required to transduce the signal from the insulin/IGF1 receptor—position 0—to the remaining elements in the pathway) does not significantly differ between both gene groups (Mann–Whitney’s $U$ test, $P = 0.816$; fig. 2A). Remarkably, genes evolving under positive selection encode proteins that have a significantly higher connectivity ($P = 0.0102$; fig. 2B), closeness ($P = 0.0042$; fig. 2C), and betweenness ($P = 0.0057$; fig. 2D) within the insulin/TOR network. These results remain significant when all kinds of interactions (i.e., not only protein–protein interactions but also metabolic and transcriptional activation interactions) are considered (table 4). Thus, the impact of positive selection on the insulin/TOR pathway genes is related to the centralities of their encoded products in the interaction network, with genes acting at the center of the network being more likely to evolve under positive selection.

| Table 1. Metrics of Network Position Taking Into Account Protein–Protein Interactions. |
|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| **Pathway**                     | **Paralogous Group**            | **Position** | **Connectivity** | **Closeness** | **Betweenness** |
| **INS**                         | **−1**                          | 1           | 0.304            | 0.000          |
| **InR**                         | **0**                           | 4           | 0.429            | 0.121          |
| **IRS**                         | **1**                           | 6           | 0.525            | 0.247          |
| **p85**                         | **2**                           | 3           | 0.375            | 0.095          |
| **p110**                        | **3**                           | 1           | 0.276            | 0.000          |
| **MELT**                        | **4**                           | 2           | 0.296            | 0.005          |
| **PKD1**                        | **4**                           | 3           | 0.429            | 0.038          |
| **PKC**                         | **5**                           | 7           | 0.553            | 0.400          |
| **TSC1**                        | **5**                           | 2           | 0.300            | 0.010          |
| **FOXO**                        | **6**                           | 2           | 0.382            | 0.081          |
| **GSK3**                        | **6**                           | 5           | 0.438            | 0.201          |
| **TSC2**                        | **6**                           | 3           | 0.396            | 0.108          |
| **EIF2BE**                      | **7**                           | 1           | 0.309            | 0.000          |
| **GYS**                         | **7**                           | 1           | 0.309            | 0.000          |
| **MYC**                         | **7**                           | 2           | 0.368            | 0.000          |
| **RHEB**                        | **7**                           | 2           | 0.356            | 0.014          |
| **TOR**                         | **8**                           | 5           | 0.477            | 0.155          |
| **4EBP**                        | **9**                           | 2           | 0.350            | 0.013          |
| **S6K**                         | **9**                           | 4           | 0.396            | 0.105          |
| **EIF4E**                       | **10**                          | 2           | 0.375            | 0.023          |
| **RPS6**                        | **10**                          | 1           | 0.288            | 0.000          |

*Position across the upstream/downstream axis of the pathway, defined as the number of steps required for signal transduction from the insulin receptor (InR).*
Table 2. List of Genomic Regions with Footprints of Positive Selection.

<table>
<thead>
<tr>
<th>Paralogous group</th>
<th>Gene</th>
<th>Geographic area</th>
<th>Method</th>
<th>#SNPs</th>
<th>Z^a</th>
<th>p^b</th>
<th>q-value^c</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRS</td>
<td>DOK1</td>
<td>MENA^c</td>
<td>iHS</td>
<td>11</td>
<td>96.2</td>
<td>2.94 x 10^{-11}</td>
<td>2.245 x 10^{-10}***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FST</td>
<td>13</td>
<td>68.1</td>
<td>1.24 x 10^{-5}</td>
<td>2.733 x 10^{-5}***</td>
</tr>
<tr>
<td>EUR</td>
<td></td>
<td></td>
<td>iHS</td>
<td>11</td>
<td>101.0</td>
<td>4.34 x 10^{-12}</td>
<td>3.985 x 10^{-11}***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FST</td>
<td>13</td>
<td>93.9</td>
<td>1.28 x 10^{-9}</td>
<td>8.428 x 10^{-10}***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>∆DAF</td>
<td>10</td>
<td>63.1</td>
<td>2.31 x 10^{-6}</td>
<td>7.510 x 10^{-7}***</td>
</tr>
<tr>
<td>EASIA</td>
<td></td>
<td></td>
<td>FST</td>
<td>13</td>
<td>99.9</td>
<td>1.33 x 10^{-10}</td>
<td>1.318 x 10^{-10}***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>∆DAF</td>
<td>11</td>
<td>79.4</td>
<td>2.00 x 10^{-8}</td>
<td>1.016 x 10^{-8}***</td>
</tr>
<tr>
<td>DOK2</td>
<td>SSAFR</td>
<td></td>
<td>FST</td>
<td>49</td>
<td>162.7</td>
<td>4.37 x 10^{-5}</td>
<td>7.544 x 10^{-6}***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>∆DAF</td>
<td>41</td>
<td>146.1</td>
<td>1.71 x 10^{-5}</td>
<td>4.80 x 10^{-6}***</td>
</tr>
<tr>
<td>MENA</td>
<td></td>
<td></td>
<td>iHS</td>
<td>45</td>
<td>213.5</td>
<td>4.81 x 10^{-12}</td>
<td>4.554 x 10^{-11}***</td>
</tr>
<tr>
<td>EUR</td>
<td></td>
<td></td>
<td>iHS</td>
<td>46</td>
<td>160.2</td>
<td>1.38 x 10^{-9}</td>
<td>2.879 x 10^{-10}***</td>
</tr>
<tr>
<td>CSASIA</td>
<td></td>
<td></td>
<td>iHS</td>
<td>44</td>
<td>357.4</td>
<td>3.67 x 10^{-34}</td>
<td>2.295 x 10^{-33}***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FST</td>
<td>49</td>
<td>170.3</td>
<td>8.29 x 10^{-6}</td>
<td>2.271 x 10^{-6}***</td>
</tr>
<tr>
<td>DOK5</td>
<td>SSAFR</td>
<td></td>
<td>FST</td>
<td>89</td>
<td>409.0</td>
<td>3.18 x 10^{-20}</td>
<td>3.977 x 10^{-19}***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>∆DAF</td>
<td>80</td>
<td>285.4</td>
<td>4.17 x 10^{-9}</td>
<td>6.361 x 10^{-9}***</td>
</tr>
<tr>
<td>MENA</td>
<td></td>
<td></td>
<td>iHS</td>
<td>88</td>
<td>419.3</td>
<td>6.75 x 10^{-22}</td>
<td>1.831 x 10^{-21}***</td>
</tr>
<tr>
<td>EUR</td>
<td></td>
<td></td>
<td>∆DAF</td>
<td>83</td>
<td>241.1</td>
<td>1.25 x 10^{-14}</td>
<td>2.244 x 10^{-14}***</td>
</tr>
<tr>
<td>CSASIA</td>
<td></td>
<td></td>
<td>∆DAF</td>
<td>79</td>
<td>237.4</td>
<td>4.47 x 10^{-8}</td>
<td>8.471 x 10^{-9}***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>iHS</td>
<td>72</td>
<td>310.6</td>
<td>2.86 x 10^{-14}</td>
<td>2.792 x 10^{-13}***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FST</td>
<td>89</td>
<td>326.8</td>
<td>7.28 x 10^{-11}</td>
<td>1.011 x 10^{-10}***</td>
</tr>
<tr>
<td>EASIA</td>
<td></td>
<td></td>
<td>∆DAF</td>
<td>71</td>
<td>270.2</td>
<td>5.02 x 10^{-10}</td>
<td>4.753 x 10^{-10}***</td>
</tr>
<tr>
<td>PKB</td>
<td>AKT1</td>
<td>MENA^c</td>
<td>iHS</td>
<td>25</td>
<td>95.0</td>
<td>1.27 x 10^{-6}</td>
<td>2.089 x 10^{-6}***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FST</td>
<td>25</td>
<td>98.7</td>
<td>4.87 x 10^{-9}</td>
<td>8.717 x 10^{-10}***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>∆DAF</td>
<td>24</td>
<td>89.4</td>
<td>2.67 x 10^{-4}</td>
<td>4.194 x 10^{-4}***</td>
</tr>
<tr>
<td>EASIA</td>
<td></td>
<td></td>
<td>FST</td>
<td>25</td>
<td>100.1</td>
<td>3.39 x 10^{-5}</td>
<td>6.12 x 10^{-6}***</td>
</tr>
<tr>
<td>AKT3</td>
<td>SSAFR</td>
<td></td>
<td>FST</td>
<td>58</td>
<td>185.4</td>
<td>4.54 x 10^{-3}</td>
<td>7.75 x 10^{-4}***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FST</td>
<td>58</td>
<td>230.1</td>
<td>7.43 x 10^{-6}</td>
<td>4.552 x 10^{-6}***</td>
</tr>
<tr>
<td>PKC</td>
<td>PRKCH</td>
<td>SSAFR</td>
<td>FST</td>
<td>132</td>
<td>508.3</td>
<td>1.19 x 10^{-17}</td>
<td>1.661 x 10^{-16}***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>∆DAF</td>
<td>116</td>
<td>365.2</td>
<td>5.33 x 10^{-8}</td>
<td>4.497 x 10^{-8}***</td>
</tr>
<tr>
<td>MENA</td>
<td></td>
<td></td>
<td>iHS</td>
<td>117</td>
<td>749.4</td>
<td>2.80 x 10^{-25}</td>
<td>5.832 x 10^{-25}***</td>
</tr>
<tr>
<td>EUR</td>
<td></td>
<td></td>
<td>∆DAF</td>
<td>116</td>
<td>369.6</td>
<td>2.30 x 10^{-8}</td>
<td>1.398 x 10^{-8}***</td>
</tr>
<tr>
<td>CSASIA</td>
<td></td>
<td></td>
<td>iHS</td>
<td>115</td>
<td>808.6</td>
<td>2.06 x 10^{-65}</td>
<td>1.287 x 10^{-66}***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>∆DAF</td>
<td>112</td>
<td>369.7</td>
<td>3.02 x 10^{-9}</td>
<td>2.224 x 10^{-9}***</td>
</tr>
<tr>
<td>PRKCI</td>
<td>OCE^c</td>
<td></td>
<td>iHS</td>
<td>125</td>
<td>430.8</td>
<td>9.12 x 10^{-12}</td>
<td>6.779 x 10^{-13}***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FST</td>
<td>29</td>
<td>135.4</td>
<td>3.95 x 10^{-8}</td>
<td>1.614 x 10^{-8}***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>∆DAF</td>
<td>24</td>
<td>103.5</td>
<td>5.94 x 10^{-6}</td>
<td>1.360 x 10^{-6}***</td>
</tr>
<tr>
<td>S6K</td>
<td>RPS6KB2</td>
<td>SSAFR</td>
<td>FST</td>
<td>20</td>
<td>112.5</td>
<td>8.09 x 10^{-9}</td>
<td>5.319 x 10^{-9}***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>∆DAF</td>
<td>17</td>
<td>91.6</td>
<td>3.47 x 10^{-7}</td>
<td>2.167 x 10^{-7}***</td>
</tr>
<tr>
<td>MENA</td>
<td></td>
<td></td>
<td>iHS</td>
<td>16</td>
<td>83.6</td>
<td>1.73 x 10^{-6}</td>
<td>5.043 x 10^{-7}***</td>
</tr>
<tr>
<td>CSASIA</td>
<td></td>
<td></td>
<td>iHS</td>
<td>15</td>
<td>116.8</td>
<td>3.54 x 10^{-12}</td>
<td>2.766 x 10^{-12}***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FST</td>
<td>20</td>
<td>100.8</td>
<td>3.74 x 10^{-7}</td>
<td>1.780 x 10^{-7}***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>∆DAF</td>
<td>15</td>
<td>68.8</td>
<td>7.00 x 10^{-5}</td>
<td>1.388 x 10^{-5}***</td>
</tr>
<tr>
<td>EASIA</td>
<td></td>
<td></td>
<td>iHS</td>
<td>17</td>
<td>118.5</td>
<td>2.78 x 10^{-11}</td>
<td>1.059 x 10^{-10}***</td>
</tr>
<tr>
<td>elf-4E</td>
<td>EIF4E</td>
<td>EASIA^c</td>
<td>iHS</td>
<td>33</td>
<td>121.9</td>
<td>3.45 x 10^{-8}</td>
<td>6.883 x 10^{-9}***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FST</td>
<td>42</td>
<td>163.3</td>
<td>5.05 x 10^{-7}</td>
<td>1.755 x 10^{-7}***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>∆DAF</td>
<td>33</td>
<td>125.5</td>
<td>1.40 x 10^{-5}</td>
<td>3.05 x 10^{-5}***</td>
</tr>
</tbody>
</table>

NOTE. N-Genes were considered to have evolved under positive selection in a given geographic area if q-value < 2.381 x 10^{-3} for the iHS test and for either the FST or ∆DAF tests (see Materials and Methods). Geographic areas for which any of these scores is significant are also displayed.

SSAFR, Sub-Saharan Africa; MENA, Middle East and North Africa; EUR, Europe; CSASIA, Central South Asia; EASIA, East Asia; OCE, Oceania.

a Fisher's combination test statistic computed from the empirical P-values obtained for SNPs within each genomic region.
b Significance associated to Z.
c Geographic areas for which the gene is considered to have evolved under positive selection.

*q-value < 2.381 x 10^{-3} (significance at 5% with Bonferroni correction for seven regions studied with three methods).

*q-value < 4.762 x 10^{-5} (significance at 1% with Bonferroni correction).

**q-value < 4.762 x 10^{-5} (significance at 1% with Bonferroni correction).
Some gene features such as expression level, expression breadth (the number of different tissues in which a gene is expressed), and the length of the encoded proteins can also affect the action of natural selection on genes (Duret and Mouchiroud 2000; Pal et al. 2001; Subramanian and Kumar 2004; Kosiol et al. 2008). Therefore, a putative correlation between these features and connectivity measures could potentially account for the observed trend. In order to rule out this possibility, we used partial correlation analysis to test for the association between positive selection (which we encoded as a binary variable) and centrality measures while controlling for the aforementioned variables simultaneously (see Hardy 1993; Cohen et al. 2003). Gene expression levels and breadths and the length of the encoded proteins were obtained from Alvarez-Ponce, Aguade et al. (2011). This analysis shows that the relationship between the patterns of positive selection and the centrality measures remains significant after controlling for these parameters (Spearman’s partial rank correlation coefficient, $\rho = 0.397$, $P = 0.005$ for connectivity; $\rho = 0.417$, $P = 0.003$ for closeness; and $\rho = 0.412$, $P = 0.004$ for betweenness). Therefore, the observed relationship between the pathway structure and the impact of positive selection is not attributable to confounding factors that influence gene evolution.

**Table 3.** Enrichment Analysis for Signals of Positive Selection among the Insulin/TOR Pathway Genes.

<table>
<thead>
<tr>
<th>Geographic Area</th>
<th>$N^a$</th>
<th>$S^b$</th>
<th>$m^c$</th>
<th>$k^d$</th>
<th>$P^e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSAFR</td>
<td>6389</td>
<td>18</td>
<td>65</td>
<td>0</td>
<td>0.168</td>
</tr>
<tr>
<td>MENA</td>
<td>6394</td>
<td>100</td>
<td>65</td>
<td>4</td>
<td>0.003**</td>
</tr>
<tr>
<td>EUR</td>
<td>6384</td>
<td>97</td>
<td>65</td>
<td>3</td>
<td>0.017*</td>
</tr>
<tr>
<td>CSASIA</td>
<td>6393</td>
<td>100</td>
<td>65</td>
<td>3</td>
<td>0.018*</td>
</tr>
<tr>
<td>EASIA</td>
<td>6365</td>
<td>151</td>
<td>65</td>
<td>1</td>
<td>0.459</td>
</tr>
<tr>
<td>AME</td>
<td>6274</td>
<td>129</td>
<td>65</td>
<td>0</td>
<td>0.743</td>
</tr>
<tr>
<td>OCE</td>
<td>6308</td>
<td>136</td>
<td>65</td>
<td>1</td>
<td>0.411</td>
</tr>
<tr>
<td>All$^f$</td>
<td>6242</td>
<td>631</td>
<td>65</td>
<td>9</td>
<td>0.116</td>
</tr>
</tbody>
</table>

*Note—SSAFR, Sub-Saharan Africa; MENA, Middle East and North Africa; EUR, Europe; CSASIA, Central South Asia; EASIA, East Asia; AME, America; OCE, Oceania.

*Genes in the genomic region-level background distribution for which a score has been computed.

*Genes evolving under positive selection in the genomic region-level background distribution.

Genes in the pathway for which a score has been computed.

*Genes evolving under positive selection in the pathway.

Statistical significance of the Hypergeometric test.

*Union of all geographic regions.

*P < 0.05; **P < 0.01.

**Fig. 2.** Comparison between genes evolving under positive selection and the remaining ones. Mean and standard error (SE) of topological measures for the nine genes under positive selection (S) and the 49 genes without signals of adaptive evolution (NS): (A) for pathway position; (B) for connectivity; (C) for closeness; and (D) for betweenness. Centrality measures were calculated taking into account the physical interactions within the insulin/TOR pathway.
Likewise, partial correlation analysis confirms the absence of association between the pathway position and the action of positive selection ($\rho = -0.203, P = 0.167$).

We also considered the effect on our observations of the threshold used to contrast whether genes evolved under positive selection (supplementary table S7, Supplementary Material online). The connectivity, closeness, and betweenness within the insulin/TOR network remain significantly higher for genes that meet the criterion to invoke positive selection when using a significance level of 0.01. If we reduce this level to 0.001, the differences of centrality measures are still higher for the genes that exhibit the footprint of positive selection, even if most of these differences do not reach significance. This can be attributed to a lack of power since we detected signatures of selection for only five of the former nine genes while using such a low significance level. Nevertheless, these results strongly support the association between the impact of positive selection on genes and the central position they occupy in the insulin/TOR pathway.

We additionally studied the differences in centrality between the nine genes evolving under positive selection and the nonsellected ones taking into account the entire interactome (Bos 

<table>
<thead>
<tr>
<th>Table 4. Relationship between the Structure of the Insulin/TOR Pathway and the Impact of Positive Selection.</th>
</tr>
</thead>
</table>
| **Network**
| **Parameter**
| **Mean for genes under positive selection**
| **Mean for genes without signal of positive selection**
| **Mann–Whitney test**
| **Partial correlation**
| **U**
| **P**
| **p**
| **P**
| ---
| Position
| 4.667
| 4.673
| 209.5
| 0.816
| -0.203
| 0.167
| Connectivity
| 5.778
| 3.612
| 103
| 0.010*
| 0.397
| 0.005**
| Closeness
| 0.500
| 0.410
| 91
| 0.004**
| 0.417
| 0.003**
| Betweenness
| 0.244
| 0.116
| 95.5
| 0.006**
| 0.412
| 0.004**
| Connectivity
| 6.889
| 4.571
| 116
| 0.024*
| 0.432
| 0.002**
| Closeness
| 0.536
| 0.455
| 101
| 0.009**
| 0.450
| 0.001**
| Betweenness
| 0.188
| 0.098
| 101.5
| 0.009**
| 0.419
| 0.003**
| Connectivity
| 35.444
| 38.087
| 186
| 0.646
| -0.132
| 0.372
| Closeness
| 0.291
| 0.294
| 179
| 0.538
| -0.170
| 0.250
| Betweenness
| 0.0115
| 0.006
| 153
| 0.228
| -0.074
| 0.618
| 

* Spearman’s partial correlation between the impact of positive selection and network parameters controlling for gene expression level and breadth and length of the encoded proteins (see Materials and Methods).

| Three different sets of interactions were used for centrality calculations: (i) protein-protein interactions within the insulin/TOR pathway; (ii) all kinds of interactions within the pathway (i.e., protein-protein, metabolic, and transcriptional activation interactions); and (iii) protein-protein interactions from the whole interactome.

**P < 0.05; **P < 0.01.

Influence of Paralogy on Positive Selection

We considered whether the probability that a gene evolved under positive selection is related to the number of paralogous copies. The number of paralogs involved in the insulin/TOR pathway (supplementary table S1, Supplementary Material online) does not significantly differ between the nine genomic regions with footprints of positive selection and the 49 remaining ones (Mann–Whitney’s U test, $P = 0.061$). Consequently, the impact of positive selection on the insulin/TOR pathway genes seems to be independent of the size of the paralogous group.

We contrasted whether genes in the same paralogous group are similarly targeted by natural selection by using a Monte Carlo method to test whether the number of pairs of paralogs where both genes show the signature of positive selection ($Y = 5$) is higher than expected at random. We observed a greater or equal $Y$ value for only 466 permutations of 10,000 ($P = 0.0466$). Therefore, when a gene evolves under positive selection, the probability that its paralogs also show signatures of adaptation increases.
Discussion

High Incidence of Positive Selection in the Insulin/TOR Signal Transduction Pathway Genes

The present analysis aims at detecting recent adaptive events (within the last 5,000–45,000 years) in human populations worldwide at genes involved in the insulin/TOR signal transduction pathway and relating these events to the structure of the pathway. To detect footprints of positive selection, we used methods devised to identify long high-frequency haplotypes (iHS) and high levels of genetic differentiation among populations (F_{ST} and ΔDAF). However, processes such as genetic drift and demographic history can also affect differentiation among populations and linkage disequilibrium patterns, making difficult to identify signals of natural selection. In particular, dramatic changes in population size such as bottlenecks and population expansions occurred during the recent human history, affecting allele frequencies (Teshima et al. 2006). Moreover, populations from America and Oceania were isolated and had small effective population sizes, and consequently must have faced dramatic genetic drift, resulting in important allele frequency fluctuations that affected the entire genome. Since nonselective processes affect the whole genome, in opposition to positive selection, which acts on specific loci, the putative effect of these factors can be ruled out by comparing the locus of interest with an empirical distribution built from all SNPs in the genome. However, it should be noted that genomic regions potentially evolving under positive selection detected using this approach might also represent false positives (Teshima et al. 2006; Pickrell et al. 2009). Therefore, we used a stringent criterion to reduce the amount of false positives raised by alternative explanations and to maximize the detection of specific footprints of positive selection. For that purpose, we combined the results from different methods based on extended haplotype homozygosity and genetic differentiation, which are expected to be differentially affected by nonselective processes, thus providing nonoverlapping false negatives. However, despite these precautions, demographic factors can increase the variance among genes, and thus, the possibility of false positives cannot be completely discarded.

The polymorphism data that we used is based on SNPs that were previously ascertained on a reduced number of HapMap samples from Africa, Europe, and East Asia. Therefore, the allele frequency spectrum observed is skewed toward high frequencies, to a different extent in the seven geographic regions. To reduce this variation among geographic regions, we removed rare variants by applying a filter according to the MAF. Moreover, the methods that we used, which are based on genetic differentiation and long-range haplotypes, reduce the effect of ascertainment bias on our results as they do not require low-frequency variants. Indeed, Romero et al. (2009) showed that this bias has a moderate effect to study population differentiation using SNP data from HGDP-CEPH; and Voight et al. (2006) claimed that the iHS method is robust to regional variation in SNP ascertainment.

Among the 65 studied genomic regions, we identified a total of nine that have potentially evolved under positive selection in human populations (table 2). The nine insulin/TOR pathway genes that fall within these genomic regions belong to five paralogous groups: IRS (insulin receptor substrate), PKB (protein kinase B), PKC (protein kinase C), S6K (S6 kinase), and elf-4e (eukaryotic translation initiation factor 4E). This relatively high incidence of adaptive evolution contrasts with a previous comparative genomics analysis showing little evidence of positive selection in the vertebrate insulin/TOR pathway (Alvarez-Ponce, Aguade et al. 2011). On the other hand, genes detected as evolving under positive selection in both analyses remarkably overlap. In vertebrates, footprints of positive selection were detected at AKT3, PRKCD, and IRS4 (the latter of which was excluded from the present analysis), which belong to the PKB, PKC, and IRS paralogous groups, respectively (Alvarez-Ponce, Aguade et al. 2011). Interestingly, our analyses revealed that AKT3 also shows a pattern of positive selection and that the PKC and IRS paralogous groups include two and three genes, respectively, with signatures of positive selection. Consistently, in the present study, we found that, in the insulin/TOR pathway, if a gene is targeted by recent positive selection, the frequency at which genes in the same paralogous group show signals of positive selection is higher than expected at random. This observation could potentially be explained by paralogous genes encoding proteins with similar structures and functions and occupying similar positions in the pathway. Taken together, these results suggest that positive selection repeatedly targets specific groups of paralogous genes within the insulin/TOR pathway.

A close examination of the gene content of the regions that evolved under positive selection, together with an analysis of the levels of linkage disequilibrium among SNPs, show that some of them also contain other genes linked to the genes of interest that might be responsible for the observed signal of positive selection (supplementary fig. S4, Supplementary Material online). Two of the studied genomic regions (DOK1 and RPS6KB2) contain a high number of additional genes with a high level of linkage disequilibrium, and thus, the signals of selection detected in these regions cannot be clearly ascribed to the insulin/TOR pathway genes. However, excluding these two genomic regions from all our analyses does not change the observed relationship between the pathway structure and the impact of positive selection (supplementary table S8, Supplementary Material online).

The distribution of footprints of positive selection among the different geographic areas shows that positive selection in the insulin/TOR pathway genes occurred mainly in Middle East and North Africa, Europe, and Central South Asia (table 3). In these regions, the proportion of genes evolving under positive selection in the insulin/TOR pathway is significantly higher than that observed in the genomic background, whereas in other geographic
regions, we observe a very low incidence of positive selection (table 3). The same tendency is observed if DOK1 and RPS6KB2 are excluded: the test remains significant in Middle East and North Africa, and it is marginally significant in Europe and in Central South Asia ( supplementary table S9, Supplementary Material online). It is worth noting that East Asia, America, and Oceania are the regions in which we detected a highest incidence of positive selection across our genomic background containing 6,394 genomic regions. This suggests that the absence of signals of positive selection in the insulin/TOR pathway genes in these three geographic areas cannot be attributed to a difficulty of detecting selective events in these areas. More problematic is the case of Sub-Saharan Africa, where we detected very few genomic regions that evolved under positive selection across the genome (18 regions of 6,394). This could be due to the general low linkage disequilibrium observed in these populations. Indeed, in these populations 1) hot spots are more uniformly distributed along the genome (1000 Genomes Project Consortium 2010); and 2) historically, recombination has impacted the genome to a greater extent because they have a higher effective population size, and they did not face any strong bottleneck like other populations. In a previous study of the nucleotide diversity of the insulin minisatellite (INS VNTR) region in another set of populations from Sub-Saharan Africa, Europe, and Asia, a pattern similar to the one observed for the other insulin/TOR pathway genes was observed, that is, potential selective events in non-African populations (Stead et al. 2003). Moreover, we found that the haplotypes that have undergone a selective sweep are shared among populations from Middle East and North Africa, Europe, and Central South Asia (see supplementary results, Supplementary Material online). Interestingly, most of those potentially shared sweeps also show significant $F_{ST}$ scores between Sub-Saharan Africa and the other regions (table 2), which can be the result of the high proportion of lower DAF values observed in Sub-Saharan Africa compared with the other areas (negative $\Delta$DAF; results not shown). Therefore, those signals of genetic differentiation could also be a consequence of the shared signatures in non-African populations, rather than evidence for selection in Africans. This suggests that positive selection acted early during the migration out of Africa (Stead et al. 2003) and that shared evolutionary history (Gutenkunst et al. 2009; Moorjani et al. 2011) accounts for the similar patterns observed in these three areas.

It has been observed that, in the United States, the prevalence of obesity and related diseases such as insulin resistance and type 2 diabetes is much lower in individuals with European and Asian ancestry (compared with Native American, Hispanic, Afro-American, and Pacific Islander groups) (Wang and Beydoun 2007; Anderson and Whitaker 2009). This observation underlines the genetic basis of these diseases, also supported by a number of heritability studies (Rankinen et al. 2006; Mathias et al. 2009; O’Rahilly 2009; Wang et al. 2009; Hinney et al. 2010). The implication of the insulin/TOR signal transduction pathway is clear since dysfunction of its components often lead to insulin resistance, diabetes, obesity, and cancer (Oldham and Hafen 2003; LeRoith et al. 2004; Taguchi and White 2008). The enrichment of adaptation events at genes involved in the insulin/TOR pathway in European, Central South Asian, and North African and Middle Eastern populations could potentially be related to the differences among populations in the prevalence of these metabolic disorders, although a direct link to specific phenotypes cannot be established yet. Thus, further analysis of the insulin/TOR pathway genes evolving under positive selection might provide insight into the etiology of these metabolic disorders.

Distribution of Positive Selection Across the Insulin/TOR Pathway

In the last decade, the availability of large-scale interaction data has allowed us to investigate the distribution of evolutionary forces across functional networks. A number of analyses have shown that there is a relationship between the structure of metabolic, regulatory, and protein–protein interaction networks and the patterns of molecular evolution of their components (for a review, see Cork and Purugganan 2004; Eanes 2011; Zera 2011). Most of these studies have focused on the levels of selective constraint acting on genes, which can be inferred from the nonsynonymous to synonymous divergence ratios, $d_{N}/d_{S}$, obtained from between-species comparisons. Values of $d_{N}/d_{S}$ near to one are expected for genes evolving neutrally, whereas values lower than one are indicative of purifying selection maintaining the sequence of the encoded proteins. Here, we considered whether the incidence of positive selection in human populations is related to the structure of the insulin/TOR pathway.

It has been observed that the $d_{N}/d_{S}$ ratios correlate with the position of genes along the upstream/downstream axis of the insulin/TOR signal transduction pathway, with downstream genes exhibiting lower $d_{N}/d_{S}$ values both in Drosophila (Alvarez-Ponce et al. 2009; Alvarez-Ponce, Guirao-Rico et al. 2011) and vertebrates (Alvarez-Ponce, Aguade et al. 2011). In addition to purifying selection, the $d_{N}/d_{S}$ ratios are also affected by the action of positive selection, and therefore, this pattern could potentially result from a higher incidence of positive selection in the upstream part of the pathway. However, in the present intraspecific analysis, we found no evidence for a polarity in the incidence of positive selection along the pathway (fig. 2A). These results contrast with a number of analyses showing that upstream genes in metabolic and transcription regulatory networks are generally subject to stronger natural selection (Rausher et al. 1999; Sharkey et al. 2005; Livingstone and Anderson 2009; Ramsay et al. 2009; Wright and Rausher 2010; Rhone et al. 2011) (but see Montanucci et al. 2011).

Whereas genes whose encoded products occupy more central positions in a protein–protein interaction network tend to evolve under stronger purifying selection (Fraser et al. 2002; Hahn and Kern 2005; Lemos et al. 2005; Montanucci et al. 2011), positive selection seems to preferentially target genes acting at the periphery of the human
protein interaction network (Kim et al. 2007). In contrast, we found that adaptive events are more frequent at genes acting at the most central positions of the insulin/TOR pathway. Indeed, proteins encoded by genes evolving under positive selection show higher connectivity, betweenness, and closeness centralities (fig. 2B–D). These most highly connected and central proteins in the pathway could potentially exert a higher influence on its function. These contrasting results indicate that the relationship between positive selection and measures of centrality deserves further consideration. Moreover, the three centrality measures calculated for the insulin/TOR pathway are highly correlated ($0.90 \leq \rho \leq 0.98$). Therefore, the observed associations between the impact of positive selection and either connectivity, closeness, or betweenness are not independent and may reflect the same underlying factors. When centrality measures were calculated from the entire interactome (Bossi and Lehner 2009) rather than hand-curated protein–protein interaction data among insulin/TOR pathway proteins (fig. 1), no difference was detected among both gene groups. Therefore, intrapathway interactions seem to have a more relevant effect on the impact of positive selection than interactions between the insulin/TOR pathway genes and the other genes. However, it should be noted that high-throughput interaction data are subject to a high proportion of false positives and false negatives (Bader et al. 2004; Deeds et al. 2006), which might obscure interactome-level analyses.

Finally, we found that genes evolving under positive selection tend to encode proteins that physically interact to each other. However, this trend vanishes when network randomization techniques that preserve the degree of each node are used. Therefore, the tendency is probably the result of genes with signatures of positive selection being involved in more interactions and hence being more likely to interact to each other. These contrasting results highlight the relevance of choosing an adequate null model for testing network properties. Further analyses of other pathways both at the interspecific and the intraspecific levels are warranted to establish whether the patterns observed in the insulin/TOR pathway are particular to this pathway or represent a general trend across the whole interactome.

**Conclusion**

We found that positive selection acted on nine of the studied genes involved in the human insulin/TOR signal transduction pathway, mostly in European, Central South Asian, and North African and Middle Eastern populations. The impact of positive selection on gene evolution depends on the position that the encoded products occupy in the pathway, with the most highly connected and central genes of the network being preferentially targeted by positive selection. These observations broaden current knowledge on how natural selection distributes across pathways within an intraspecific framework, more particularly considering short periods of population-specific adaptation.

**Supplementary Material**

Supplementary results and figures S1–S4 and tables S1–S9 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

**Acknowledgments**

Bioinformatics services were kindly provided by the Genomic Diversity node, Spanish Bioinformatic Institute. This work was funded by grants BFU2010-19443 (subprogram BMC) awarded by Ministerio de Educación y Ciencia (Spain) and the Direcció General de Recerca, Generalitat de Catalunya (Grup de Recerca Consolidat 2009 SGR 1101). P.L. is supported by a PhD fellowship from “Acción Estratégica de Salud, en el marco del Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica 2008–2011” from Instituto de Salud Carlos III.

**References**


