[Frontiers in Bioscience 11, 380-388, January 1, 2006]

Co-evolutionary analysis of insulin/insulin like growth factor 1 signal pathway in vertebrate species

Tonghai Dou, Chaoneng Ji, Shaohua Gu, Jiaxi Xu, Jian Xu, Kang Ying, Yi Xie, and Yumin Mao

State Key Laboratory of Genetic Engineering, Institute of Genetics, School of Life Science, Fudan University, Shanghai 200433 PR, China

TABLE OF CONTENTS

1. ABSTRACT

Insulin/insulin like growth factor 1 signaling pathway plays an essential role on the regulation of variant, interrelated and fundamental processes, such as metabolism, growth, reproduction and aging. This pathway is highly conserved during evolution. We analyzed eight vertebrate species' insulin/insulin like growth factor 1 pathway and examined the co-evolutionary relationship between different protein components by quantifying similarity of phylogenetic trees. The collected species include *Takifugu rubripes*, *Danio rerio*, *Xenopus laevis*, *Gallus gallus*, *Mus musculus*, *Rattous norvegicus*, *Pan troglodytes* and *Homo sapiens*. Results show that interacting proteins in this pathway share highly co-evolutionary relationship while contrarily the proteins without interaction have low co-evolutionary relationship. We also predict some receptor and ligand partners enjoy highly binding potential in corresponding species' pathway.

2. INTRODUCTION

Insulin/IGF-I (insulin like growth factor 1) signaling pathway is an evolutionarily conserved pathway that plays an essential role in the storage and release of energy during feeding and fasting a large portion of somatic growth. Recent researches also indicate this pathway related to the life span (1- 5). The insulin and IGF-I signaling cascade, commonly consisting of insulin, IGF-I, IGF-II, insulin receptor, IGF-I receptor, insulin receptor substrates (IRSs) and other substrate proteins, exists in vertebrate and has homologous pathways in invertebrates. The Insulin and IGF family contains insulin, IGF-I and IGF-II. Insulin is widely known for its regulating blood glucose (4). However, insulin and IGF-I have a broader role in promoting somatic growth, tuning tissue development and survival (6). In general, these ligands enjoy high level of binding affinity with their corresponding receptors. Moreover, 'cross-binding' does

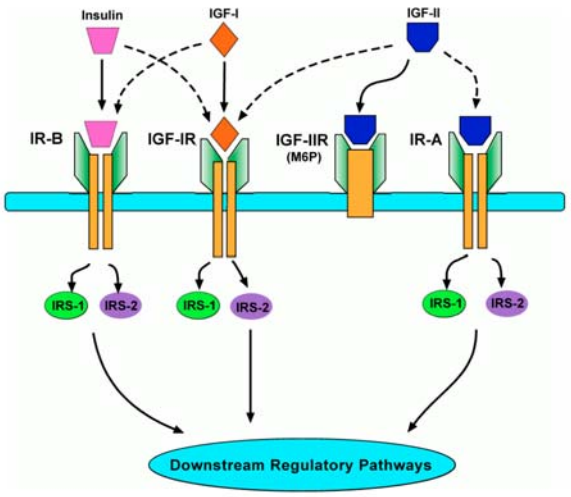

Figure 1. The interacting relationship of insulin/IGF superfamily ligands, receptors and downstream molecules. Insulin, IGF-I, IGF-II all have high level of affinity with their corresponding receptors (shown in real line). On the other hand, these ligand can 'cross-bind' with other receptors as well (shown in dashed line), but with lower affinity. IGF-II is considered related to the regulation of body growth and embryonic development. Its specific receptor, type II IGF receptor (IGF-II receptor), is identical to the cation-independent mannose-6-phosphate (M6P) receptor. Meanwhile, IGF-II can also bind to two other homologous proteins, IGF-I receptor and one isomer of insulin receptor IR A (7).

exist as well (Figure 1). Through different expression level in organs, as well as phosphorylation of different substrates, these receptors regulate various signaling pathways such as MAP and PI3K pathways (7). IGF-II receptor has no intracellular domain, which is different from IR and IGF-IR, and is responsible for disassembling IGF-II. Substrates of IR/IGF-IR include insulin receptor substrates (IRS) and some other substrate protein as SHC, APS and GAB1/2, DOCK1/2 etc., among which IRS serves as the crucial one. In spite of various kinds of IRS, recent research suggests that many insulin responses are largely regulated through two IRS proteins, IRS1 and IRS2 (6). IRS1 mediates the effects of insulin and IGF-I on somatic cell growth, whereas IRS2 is essential for nutrient homeostasis (8, 9). In vertebrate, the IRS1 and IRS2 are relatively conserved.

Insulin/IGF-I signal pathway is a cross species conserved pathway. From yeast to mammals, all contain a similar signal cascade (10). Between the invertebrate and vertebrate there exists noticeable difference. For example, *C. elegans* does not have presently detectable IRS homolog (11), while *D. melanogaster* has several insulin like proteins (ILPs) (12). However, in vertebrates this pathway is well conserved and widely distributed. Zebrafish, *Takifugu rubripes*, *Xenopus laevis*, Jungle Fowl, Norway rat, mouse, chimpanzee and human all contain this signal pathway: the insulin and IGFI bind to their corresponding receptors; IR and IGFIR phosphorylate the IRSs and other substrates (Figure 2). Although there are four mammalian IRSs (IRS1-4), research indicates the IRS1 and IRS2 play essential roles in the signal pathway (6). Furthermore, in human and chimpanzee the presumable IRS3 genomic region may lose transcript function due to the stop codon insertion (13). Difference of this signal pathway also exists in fish and bird. In fish, this is probably due to the genome duplication (14, 15). Zebrafish and fugu have double insulin, IR and IGF-IR genes, while large-scale-sequencing shows that there may be several paralogous IRSs in fish and Chicken genome. The evolutionary analysis indicates that duplication of the original ligand created the present three ligands (16), and their corresponding receptors experienced duplication as well. As to the emerging of downstream molecules, they are probably introduced to this pathway subsequently, which is substantiated by comparing the human pathway with the nematode's (Figure 2). In general, vertebrate's insulin/IGF-I signal pathways all contain INS, IGF-I, IGF-II, IR, IGF-I R and IRS protein homologs (10).

Prediction and identification of protein-protein interaction is always the focus of biological research. Experimental approaches for the determination of interaction proteins have been rapidly developed since the 1980s, especially through the yeast-two-hybrid protocol scanning (17). At the same time, many computational approaches to predict interacting partners have been developed since more and more genome and protein sequences are available, such as phylogenetic profiles; conservation of gene neighboring; correlated mutations and similarity of phylogenetic trees (18). Among these methods, the similarity of phylogenetic tree method is a convenient and powerful one. It is based on the hypothesis that the interacting or functionally related proteins may have similar selective pressure, resulting in the similarity of the evolutionary pattern (such as evolutionary tree). Using statistical method on phylogenetic tree data, Goh et al first quantified this method based on the co-evolution of ligand-receptor family in evolving organisms (19). And after that they developed the method to infer the co-evolutionary relationship of binding proteins using the correlation coefficient between sequence similarity distance matrices (20) Besides, Pazos and Valencia statistically demonstrated the relation and interaction in large sets of interacting proteins in *E.Coli* using similar method (21). These results indicate the interacting protein pairs' co-evolution is distinct and potentially interacting pairs' correlation values may be no less than 0.7-0.8 (19, 20, 21).

Although Fryxell mentioned the similarity of phylogenetic tree between insulin and IR family (22), quantitative analysis of co-evolutionary relationship in insulin/IGF-I signal pathway hasn't been reported. In this study, we collect *Takifugu rubripes*, *Danio rerio*, *Xenopus laevis*, *Gallus gallus*, *Mus musculus*, *Rattous norvegicus*, *Pan troglodytes*, *Homo sapiens*'s corresponding protein sequences in this pathway. Based on multiple sequence alignment and Goh's method (19, 20), we computed the correlation coefficient of different protein pairs in the insulin/IGF-I signal pathway. Our result shows that the correlation value in known interacting protein pairs is much

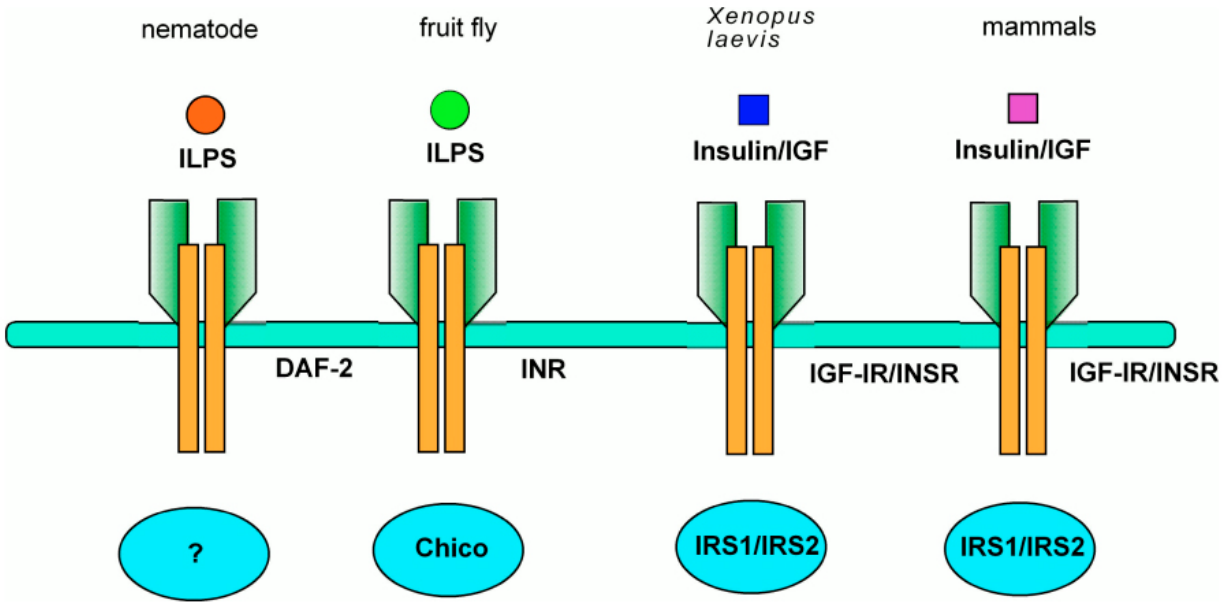

Figure 2. Comparison of insulin/IGF pathways in different animals. Noticeable differences are detected between the invertebrate and vertebrate. Instead of insulin/IGF, *C. elegans* and *D. melanogaster* contains several insulin like proteins (ILPs), responsible for various biological functions. And their downstream molecules differ from vertebrates as well. However, in vertebrates the insulin/IGF pathway is well conserved and widely distributed, comprising of insulin/IGF ligand, receptors and IRSs (mainly IRS1 and IRS2).

higher than the negative control pairs, and different binding affinity pairs shows different correlation value. Furthermore, we also indicate some potentially interacting partners from homologous genes in some species.

3. MATERIAL AND METHODS

3.1. Molecular databases and sequence analysis

Genomic database maintained by the NCBI (http://www.ncbi.nlm.nih.gov/), Ensemble (http://www.ensembl.org/), the zebrafish genome initiative (http://zfblasta.tch.harvard.edu), the *Danio rerio* sequencing project (http://www.sanger.ac.uk/Projects/D_rerio), and Project of chimpanzee genome and Gallus genome in Washington University Genome Sequencing Center (http://www.genome.wustl.edu/projects/) were searched for the proteins in the insulin/IGF-I signal pathway. The culled sequences contain the homologs of insulin, IGF-I, IGF-II, IR, IGF-IR, insulin receptor substrates (IRSs) and IGF-IIR. Besides, we also choose a negative control gene cytochrome P450, family 17, Cyp17a1 that is not directly related to above-mentioned proteins. Gene prediction and identification are based on genomic gene annotation system provided by Ensembl (23) and verified by NCBI GenomeScan (24). Duplicated gene are collected and put into further analysis. Limited by the available protein sequence of IGF-I protein of *Gallus gallus* that contains only 153 residues, we use similar transcript variant in other species. Besides NCBI homologen and Ensembl ortholog gene analysis, all sequences are BLAST against protein sequences, gene sequences and EST sequences in different databases (Table 1).

3.2. Co-evolutionary correlation analysis

Multiple protein sequence alignment is performed by ClustalW (25) and confirmed by VectorNTI 6.0 (InforMax. Inc.) and PHYLIP (26). Evolutional pair-wise distance matrices are constructed by MEGA2 (27) based on multiple alignment results. So nine genes' matrices were constructed including insulin, IGF-I, IGF-II, IR, IGF-IR, IRS1, IRS2, and Cyp17a1. Potential paralogs and homologs' distance matrices are constructed respectively. Each matrix contains eight species and is an 8×8 matrix. Due to lacking *Xenopus laevis* sequence, the IGF-IIR matrix only contains seven species. The correlation value of each protein pair, such as insulin and IR, is generated using statistical method based on the insulin matrix and IR matrix. For an X-Y protein pairs, the linear correlation coefficient r [Pearson's correlation coefficient, 28] of this protein pair is calculated according to the standard equation:

$$r = \frac{\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\left(X_{ij}-\overline{X}\right)\left(Y_{ij}-\overline{Y}\right)}{\sqrt{\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\left(X_{ij}-\overline{X}\right)^2}\sqrt{\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\left(Y_{ij}-\overline{Y}\right)^2}} \quad \text{(equation 1)}$$

Where X_{ij} represents the distance of i species to j species in former protein matrix, $\overline{X}$ represents the mean of all Xij values, Y_{ij} represents the distance of i species to j species in latter protein matrix, $\overline{Y}$ represents the mean of all Yij values, and N represents the species number of the matrices.

Table 1. Sequences culled in co-evolutionary analysis of insulin/IGF-I signal pathway

	Homo sapiens	***Pan troglodytes***[1]	***Mus musculus***	***Rattus norvegicus***
Insulin	NM_000207	X61089	NM_008386, NM_008387	NM_019129, NM_019130
IGF-I	NM_000618	ENSPTRG00000005362	NM_184052	AAA41215
IGF-II	NM_000612	ENSPTRG00000003170	NM_010514	NM_031511
IR	NM_000208	ENSPTRG00000010386	NM_010568	NM_017071
IGF-I R	NM_000875	ENSPTRG00000007490, NM_000875[a]	NM_010513	NM_052807
IGF-II R	NM_000876	ENSPTRG00000018764	NM_010515	NM_012756
IRSs	NM_005544, NM_003749, NM_003604	ENSPTRG00000012989, ENSPTRG00000006030, ENSPTRG00000022172	NM_010570, XM_357863, NM_010571, NM_010572	NM_012969, AAB35238, ENSRNOG00000023509, NM_032074, XM_235721
Cypt17a1	NM_000102	ENSPTRG00000002900	NM_007809	NM_012753
	Gallus gallus	***Xenopus laevis***	***Danio rerio***	***Takifugu rubripes***
Insulin	NM_205222	M24443, M24442	NM_131056, ENSDARG00000011006	SINFRUG00000124023, SINFRUG00000143863
IGF-I	NM_001004384	M29857	NM_131825	SINFRUG00000140885
IGF-II	AAB46818	AY050645	NM_131433, NM_001001815	AL021880
IR	XM_418250	AJ132556	ENSDARG00000011948, ENSDARG00000021712	SINFRUG00000136186, SINFRUG00000147733
IGF-I R	NM_205032	AF055980	NM_152968, NM_152969	SINFRUG00000141342, SINFRUG00000144438
IGF-II R	NM_204970		ENSDARG00000006094	SINFRUG00000138504
IRSs	U43502, ENSGALG00000008107, ENSGALG00000004959, ENSGALG00000016839, ENSGALG00000003823	U27842, BC072768, AF297960	ENSDARG00000007351, ENSDARG00000018236, ENSDARG00000019776, NM_200315, ENSDARG00000009341, ENSDARG00000015855	SINFRUG00000152504, SINFRUG00000149420, SINFRUG00000121907, SINFRUG00000139560, SINFRUG00000158028
Cypt17a1	NM_001001901	AF325435	NM_212806	SINFRUG00000155507

[1] Chimpanzee genome sequence isn't completely available, some protein sequences' information are based on Homo sapiens, [2] Each sequence is verified in different databases.

The correlation coefficient r' of this protein pair in a given species k ($1 \le k \le N$) is calculated according to the standard equation:

$$r' = \frac{\sum_{i=1}^{N}\left(X_{ki} - \overline{X}_k\right)\left(Y_{ki} - \overline{Y}_k\right)}{\sqrt{\sum_{i=1}^{N}\left(X_{ki} - \overline{X}_k\right)^2}\sqrt{\sum_{i=1}^{N}\left(Y_{ki} - \overline{Y}_k\right)^2}} \quad \text{(equation 2)}$$

In which Xki and Yki represent two row vectors of k species in matrix X and matrix Y, $\overline{X}_k$ and $\overline{Y}_k$ represent the mean of all distance in the row vector Xk and Yk.

In each protein pair calculation, we employ the presumably false binding pairs and proteins with Cyp17a1 pairs as negative control.

4. RESULTS

4.1. Interaction between IR, IGF-IR and ligands

Ligands and receptors we study here exhibit cross interaction to some extent (Figure 1). On the other hand, IR and IGF-IR can regulate their downstream molecules. To get a clear picture about the evolutionary relation of ligand, IR, and IGF-IR, we calculated their correlation coefficient. Our data show ligands share high correlation value with these two receptors, which are higher than 0.8. Besides, the cross protein pairs' calculation shows that insulin has a little higher co-evolutionary relationship with IR than IGF-I and IGF-II. So does IGF-I for IGF-IR, but IGF-II shows relatively lower correlation coefficient with these two receptors (Figure 3).

4.2. Co-evolution of protein pairs in insulin/IGF-I pathway

By multiple alignment and orthologous gene analysis, we selected insulin, IGF-I, IGF-II, IR, IGF-IR, IRS1, IRS2 and cypt17a1 proteins in eight species and IGF-II receptor in seven species into calculation. Particular values of protein pairs are shown in Table 2. Through the calculation, we find the interaction protein pairs gained higher correlation value (all higher than 0.8), such as insulin-IR, IGFI-IGF-IR, IR-IRS1, IGF-IR-IRS1. And the correlation value of some related protein pairs, such as insulin-IRS1, decrease a little (0.7-0.8) (Table 2). All the negative protein pairs gained quite lower correlation value (less than 0.7).

Co-evolution analysis of receptors and their substrates shows that IR and IGF-IR have close relationship with IRS1 and IRS2 protein (higher than 0.8, Table 2). But the value is lower compared to IR and IGF-IR with their ligands. This may be partially due to the fact that IRS1 and IRS2 interact with variant receptors and downstream signal molecules, which in turn limits their co-evolution relation with IR and IGF-IR. Without intracellular tyrosine domain, IGF-II R can't interact with IRS proteins; and this is strongly supported by its low correlation values with IRS proteins (Table 2). Moreover, we also perform the analysis on the protein pairs of insulin, IGF-I and IGF-II to IRS proteins. The correlation values are much lower than those of ligand-receptor pairs (Delta $\overline{r}$ =0.16) and receptor-substrate pairs (Delta $\overline{r}$ =0.08) (Figure 4).

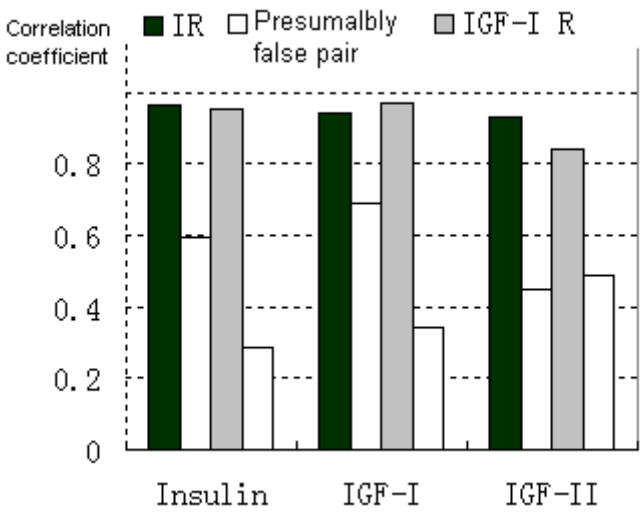

Figure 3. Correlation coefficient of insulin/IGF superfamily ligands with their corresponding receptors. The figure shows the correlation coefficient for the ligands with IR and IGF-IR. The correlation values of known interaction pairs are remarkably higher than the presumably false pairs. For IR, the correlation value with insulin is slightly higher than IGF-I. And the IR-IGF-II pair is respectively lower than the other two ligand-receptor pairs: insulin-IR, IGF-I-IR. Detail data are shown in Table 2.

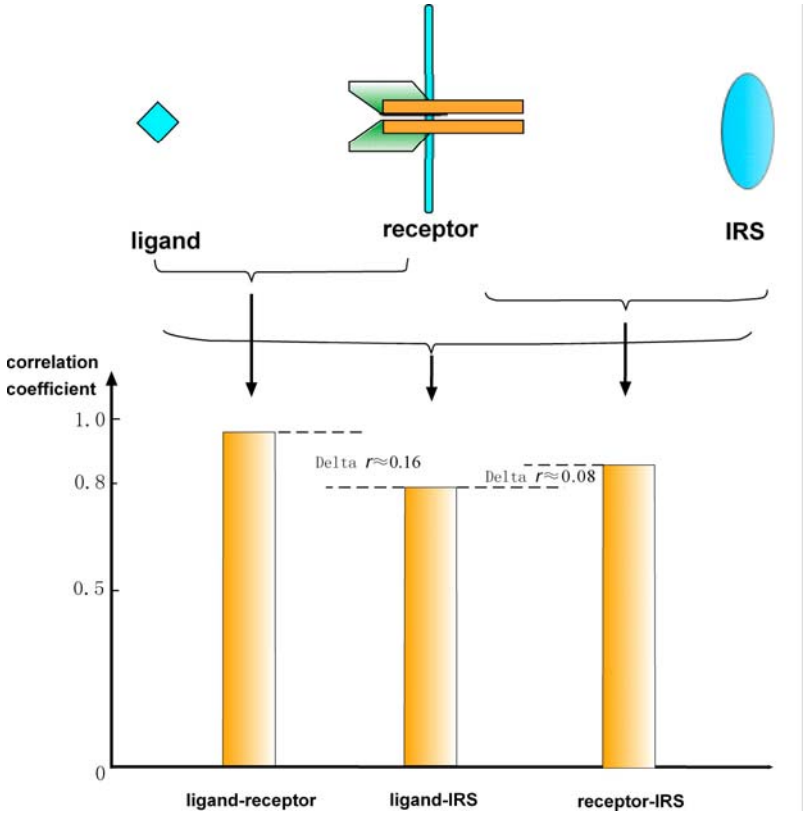

Figure 4. Correlation values of different kinds protein pairs. First column (left column) represents the average correlation coefficient ($\bar{r}$ =0.96) of ligand-receptor pairs, the right column illustrates the average correlation coefficient ($\bar{r}$ =0.87) between IRSs and receptors. They both show considerably high level of correlation. The column in the middle means the average correlation coefficient ($\bar{r}$ =0.79) between ligands and IRSs. And it can be observed that this correlation values are lower than those of ligand-receptor pairs (Delta $\bar{r}$ =0.16) and receptor-substrate pairs (Delta $\bar{r}$ =0.08).

Table 2. Correlation coefficient of different protein pairs

	IR	IGF-I receptor	IGF-II receptor [1]	IRS-1	IRS-2	Cypt17a1
insulin	0.9651	0.9522	0.9679 [1]	0.7964	0.7801	0.7468
IGF-I	0.9427	0.9702	0.9643 [1]	0.8483	0.7479	0.6724
IGF-II	0.9290	0.8399	0.9595 [1]	0.6332	0.8384	0.7174
IR				0.8066	0.8904	0.6446
IGF-I receptor				0.9227	0.8625	0.5732
IGF-II receptor [1]				0.7904 [1]	0.7871 [1]	0.7432 [1]
IRS-1						0.3637
IRS-2						0.5264

[1] Correlation coefficient using 7 species, Negative control values are list in the end line of Cyp17a1 in the table.

Table 3. Correlation value of homologous protein pairs

Part One		**Fugu Insulin1 SINFRUG00000124023**	**Fugu insulin2 SINFRUG00000143863**
Fugu insulin Receptor	SINFRUG00000136186	0.9671	0.9647
	SINFRUG00000147733	0.9720	0.9878
Fugu IGF-I receptor	SINFRUG00000141342	0.9719	0.9698
	SINFRUG00000144438	0.9246	0.9566
		Zebrafish Insulin NM_131056	Zebrafish Insulin2 ENSDARG00000011006
Zebrafish insulin receptor	ENSDARG00000011948	0.9633	0.9739
	ENSDARG00000021712	0.9679	0.9930
Zebrafish IGF-I Receptor	NM_152968	0.9414	0.9692
	NM_152969	0.9265	0.9680
Part Two		Gallus IR XM_418250	Gallus IGF-I R NM_205032
Gallus IRS1 homolog protein	**U43502**[1]	**0.8629**	**0.9389**
	ENSGALG00000004959	0.5310	0.3492
Gallus IRS2 homolog protein	ENSGALG00000008107	0.6149	0.6373
	ENSGALG00000016839[1]	**0.8945**	**0.7507**
		Fugu IR SINFRUG00000136186,	Fugu IGF-I R SINFRUG00000147733
Fugu IRS1 homolog protein	SINFRUG00000149420	0.9096	0.9469
	SINFRUG00000121907	0.8960	0.9374
	SINFRUG00000158028[1]	**0.9210**	**0.9531**
Fugu IRS2 homolog protein	SINFRUG00000152504	0.9104	0.8969
	SINFRUG00000158028[1]	**0.9175**	**0.9028**
		Zebrafish IR ENSDARG00000011948	Zebrafish IGF-I R NM_152968
Zebrafish IRS1 homolog protein	**ENSDARG00000019776**[1]	**0.9210**	**0.9531**
	ENSDARG00000007351	0.8461	0.7331
Zebrafish IRS2 homolog protein	**NM_200315**[1]	**0.9395**	**0.9092**
	ENSDARG00000018236	0.8381	0.8338

[1] Potential orthologous genes are in bold and underlined.

These results indicate the interacting proteins in insulin/IGF-I signal pathway may have similar evolutionary pressure and thus gain close co-evolutionary relationship. And the neighboring protein pairs (directly interacting) in the pathway are more likely to have higher co-evolutionary relationship than distant protein pairs (not directly interacting and with an extra steps apart).

4.3. Duplicated or homologous gene analysis

Zebrafish and fugu genome contain two potential paralogous insulin, IR and IGF-IR genes, probably due to the ancient genome duplication. When constructing different pair-wise distance matrices for each possible duplicated gene pairs, we found each protein pair has high correlation efficient (higher than 0.9). As expected, small differences exist, especially in the two insulin genes of zebrafish. In which zebrafish insulin2 (ENSDARG00000011006) share higher correlation value with four receptor proteins than the insulin gene (NM_131056) (Table 3, Part One).

Although mammals have four IRS proteins, we only found IRS1 and IRS2 have presumable homologous genes in *Gallus gallus*, *Xenopus laevis*, *Danio rerio* and *Takifugu rubripes*. Multiple sequence alignment shows that *Gallus gallus*, *Danio rerio* and *Takifugu rubripes* all have several IRS protein homologs. We constructed pair-wise distance matrix for each presumable homologous IRS protein separately, and analyzed the correlation value with IR and IGF-IR. In *G. gallus*, sequences of U43502 and

ENSGALG00000016839 have remarkable higher correlation values with the receptors than other homologous IRS proteins. And in zebrafish and fugu, each analyzed data set is list in Part Two of Table 3. The potential orthologous genes with highest correlation value are in bold and underlined. Negative control are performed and presumable false pairs' value is lower than 0.6 (data not shown).

5. DISCUSSION

Insulin/IGF-I signal pathway has important function to organism, and the whole system is conserved in phylogenetic development. The biological processes, which regulated by the pathway in different species, are rather similar, such as oxidative pressure, metabolism regulation, food utilization and life cycle (5, 10). Similar evolutionary pressure on this pathway provides the foundation for co-evolutionary analysis, while various species genome sequencing provides necessary cross-species sequence information. In Goh et al and Pazos's researches, the interaction protein pairs' correlation values are nearly higher than 0.8 (19, 20, 21). In our analysis, we found that the protein pairs in insulin/IGF-I signal pathway with clearly interaction relationship share high correlation value (higher than 0.8), such as insulin-IR, insulin-IGF-IR and IR-IRS1. As for those protein pairs which may not interact, such as IGF-IIR-IRS1, the correlation value is relatively lower (less than 0.7). And all proteins with presumably negative control protein CYP17a1 share the lowest value (less than 0.7) (Table 2). This indicates that the co-evolutionary analysis for insulin/IGF-I signaling pathway basically represents the biological interacting relation, which in turn can be used as a useful tool to predict protein interaction relation.

In insulin/IGF1 signaling pathway, ligands and receptors share cross-binding relationship, and bind their corresponding partner with higher affinity (compared with other 'cross-binding partners they can interact) (Figure 1). Our data show these three ligands share high correlation value with their specific receptors, which are higher than 0.9. The cross protein pairs' calculation shows that insulin has a little higher co-evolutionary relationship with IR than IGF-I and IGF-II (Figure 3).Also similar result can be seen on protein pairs of IGF-I receptor with these ligands. These data may partly represent the binding affinity order, though the discrimination is comparatively low. Interestingly, IR and IGF-IR can form as functional hybrids (29), which also reflect the resembling evolution pressure of these two receptors.

We also obtained higher coefficient value of IGF-II with IR than with IGF-IR, this may reveal the tightly interaction of IGF-II with IR-A during embryogenesis which mentioned above (Figure 1, Table 2). Because of *Xenopus laevis* IGF-II receptor hasn't been identified, the analysis of IGF-IIR with other proteins is carried out upon other seven species. The co-evolutionary relationship of IGF-IIR with three ligands above shows insulin has stronger interaction with IGF-IIR than IGF-I and IGF-II (Table 2). Although insulin hasn't direct interaction with IGF-IIR, together with IGF-II, they are all imprinted genes in mammals (30, 31). But the imprinting of IGF-II and IGF-IIR seems doesn't exist in chickens (31, 32). Furthermore, in chicken and *Xenopus Laevis*, experiments in vitro shows IGF-II couldn't bind to IGF-II receptor (33). And researches suggest that IGF-IIR may be recruited from other pathways during evolution of higher animals (34), which probably causes the overall lower correlation coefficient of IGF-II and IGF-IIR. For further analysis, we add insulin, IGF-I, IGF-II and IGF-IIR sequences from two mammalian species (*Sus scrofa* and *Bos taurus*) into calculation. At the same time, two fish and chicken's sequences are removed. Result shows the protein pair's correlation values are aligned in the order of IGF-II-IGF-IIR (0.9865), insulin-IGF-IIR (0.9810), IGF-I-IGF-IIR (0.8810), which is from high to low. Interestingly, though IGF-II has the highest values, insulin also shares nearly similar score. Whether this phenomenon due to imprinted genes (insulin, IGF-II, IGF-IIR) have similar evolution pattern need more confirmation.

Gene duplication is a common phenomenon during evolution. New paralogous gene will arise via multiple ways, such as genome duplication, retrotransposon insertion. In the insulin/IGF-I signal pathway, we find that zebrafish and fugu both contain two insulin, IR and IGF-IR genes. This probably due to genome scale duplication events, which might have occurred in the ancestor of fish (14, 15). Our result shows in fugu, two insulin share high correlation value with two IR and IGF-IR protein. Fugu insulin 2 (SINFRUG00000143863) shares higher co-evolutionary relationship with fugu IR 2 (SINFRUG00000147733) and fugu IGF-IR 2 (SINFRUG00000144438). Fugu insulin 1 (SINFRUG00000124023) share higher correlation value with the other two receptors. Fugu IR 1 (SINFRUG00000136186) and fugu IGF-IR 1 (SINFRUG00000141342). In zebrafish, interestingly, the paralogous insulin gene zebrafish insulin2 (ENSDARG00000011006) share higher correlation value with four receptor proteins than the insulin gene (NM_131056) (Table 3). Whether the result predicts these protein pairs have different binding affinity still need to be confirmed. And recent research indicates in fish, two paralogous insulin genes may have different function and zebrafish insulin2 may have more ubiquitous expression pattern than insulin1 (35). All these, to a certain extent, reflect the divergence of the co-evolutionary relationship of fish paralogous genes. As to rat and mouse, they also contain duplicated insulin gene, but correlation analysis result shows they have nearly same co-evolutionary relationship with IR and IGF-IR (data not shown). As a result, though two paralogous insulin genes in rat and mouse have diverse function (36, 37), their slight evolutionary divergence may be difficult to distinguish by this method.

During IRS protein analysis, we found that *Gallus gallus* and fish all contain several putative IRS proteins. Multiple alignment shows chicken and zebrafish contain double homolog genes of IRS1 and IRS2 protein (Table.3, Part Two). Ensemble ortholog gene analysis indicates these

two genes are orthologs of IRS1 and IRS2. Through co-evolutionary analysis, we obtain the coefficients with obvious difference. And the ones, which have higher correlation value, may be the potential orthologs (Bold and underlined in Table 3, Part Two). And in fugu, multiple sequence alignment shows its genome doesn't contain obvious IRS1 and IRS2 protein homolog genes. So we select all IRS proteins for calculation, result shows sequence SINFRUG00000158028 has higher co-evolutionary value in the IRS1 and IRS2 protein analysis. Whether it is due to the disruption caused by the great sequence difference between Fugu IRS orthologs and other species' IRSs needs further verification.

It should be emphasized that such duplication of ligand and receptor also exist in some other signaling pathways. For instance, in the studies on TGF-beta superfamily, annotating gene functions by comparing phylogenetic trees directly has been used (38). And it can be imagined this abovementioned quantitative method will facilitate such signaling pathway analysis. Moreover, because co-evolution is wide observed in hormone regulatory pathways such as luteinizing hormone/follicle-stimulating hormone, vasopressin/oxytocin and thyrotropin (39-41), the method employed in our research can be very helpful in predicting the co-evolutionary interacting partner of orphan receptors and ligands. Last but not least, this quantitative method can assist to annotating gene functions in species whose genome sequence have been available.

Using quantitative similarity of phylogenetic tree method we analyze the co-evolutionary relationship of eight species' insulin/IGF-I signal pathway. The directly interacting protein pairs share higher correlation values than non-directly interacting proteins pairs. Also we examined the paralogous and homologous proteins and predict the potential orthologous proteins based on closest coevolution relationship. Further researches are expected such as binding affinity of zebrafish insulin1 and insulin2 with different receptors.

6. ACKNOWLEDGMENT

We thank Dr. Hai Wu, Dr. Yi Zhang, and Dr. Jielin Yang for helpful suggestions regarding this work, and M.Sc. Yi Chen for the design of the analysis program. Financial support was provided by the 863 project of PR China (grant number: 2001AA221181).

7. REFERENCES

1. Bartke A, Chandrashekar V, Dominici F, Turyn D, Kinney B, R. Steger & J. J. Kopchick: Insulin-like growth factor 1 (IGF-1) and aging: controversies and new insights. *Biogerontology* 4, 1-8 (2003)
2. Clancy D. J, Gems D, Harshman L. G, Oldham S, Stocker H, Hafen E, S. J. Leevers & L. Partridge: Extension of life-span by loss of CHICO, a Drosophila insulin receptor substrate protein. *Science* 292, 104-106 (2001)
3. D. Gems & L. Partridge: Insulin/IGF signalling and ageing: seeing the bigger picture. *Curr Opin Genet Dev* 11, 287-292 (2001)
4. Tatar M, A. Bartke & A. Antebi: The endocrine regulation of aging by insulin-like signals. *Science* 299, 1346-1351 (2003)
5. Rincon M, Muzumdar R, G. Atzmon & N. Barzilai: The paradox of the insulin/IGF-1 signaling pathway in longevity. *Mech Ageing Dev* 125, 397-403 (2004)
6. M. F. White: IRS proteins and the common path to diabetes. *Am J Physiol Endocrinol Metab* 283, E413-422 (2002)
7. Nakae J, Y. Kido & D. Accili: Distinct and overlapping functions of insulin and IGF-I receptors. *Endocr Rev* 22, 818-835 (2001)
8. Withers D. J, Gutierrez J. S, Towery H, Burks D. J, Ren J. M, Previs S, Zhang Y, Bernal D, Pons S, Shulman G. I, S. Bonner-Weir & M. F. White: Disruption of IRS-2 causes type 2 diabetes in mice. *Nature* 391, 900-904 (1998)
9. D. J. Withers & M. White: The insulin signaling system--a common link in the pathogenesis of type 2 diabetes. *Endocrinology* 141, 1917-1921 (2000)
10. Barbieri M, Bonafe M, C. Franceschi & G. Paolisso: Insulin/IGF-I-signaling pathway: an evolutionarily conserved mechanism of longevity from yeast to humans. *Am J Physiol Endocrinol Metab* 285, E1064-1071 (2003)
11. H. A. Tissenbaum & G. Ruvkun: An insulin-like signaling pathway affects both longevity and reproduction in Caenorhabditis elegans. *Genetics* 148, 703-717 (1998)
12. Brogiolo W, Stocker H, Ikeya T, Rintelen F, R. Fernandez & E. Hafen: An evolutionarily conserved function of the Drosophila insulin receptor and insulin-like peptides in growth control. *Curr Biol* 11, 213-221 (2001)
13. Bjornholm M, He A. R, Attersand A, Lake S, Liu S. C, Lienhard G. E, Taylor S, P. Arner & J. R. Zierath: Absence of functional insulin receptor substrate-3 (IRS-3) gene in humans. *Diabetologia* 45, 1697-1702 (2002)
14. Amores A, Force A, Yan Y. L, Joly L, Amemiya C, Fritz A, Ho R. K, Langeland J, Prince V, Wang Y. L, Westerfield M, M. Ekker & J. H. Postlethwait: Zebrafish hox clusters and vertebrate genome evolution. *Science* 282, 1711-1714 (1998)
15. Taylor J. S, Braasch I, Frickey T, A. Meyer & Y. Van de Peer: Genome duplication, a trait shared by 22000 species of ray-finned fish. *Genome Res* 13, 382-390 (2003)
16. Patton S. J, G. N. Luke & P. W. Holland: Complex history of a chromosomal paralogy region: insights from amphioxus aromatic amino acid hydroxylase genes and insulin-related genes. *Mol Biol Evol* 15, 1373-1380 (1998)
17. S. Fields & O. Song: A novel genetic system to detect protein-protein interactions. *Nature* 340, 245-246 (1989)
18. Xia Y, Yu H, Jansen R, Seringhaus M, Baxter S, Greenbaum D, H. Zhao & M. Gerstein: Analyzing cellular biochemistry in terms of molecular networks. *Annu Rev Biochem* 73, 1051-1087 (2004)
19. Goh C. S, Bogan A. A, Joachimiak M, D. Walther & F. E. Cohen: Co-evolution of proteins with their interaction partners. *J Mol Biol* 299, 283-293 (2000)
20. C. S. Goh & F. E. Cohen: Co-evolutionary analysis reveals insights into protein-protein interactions. *J Mol Biol* 324, 177-192 (2002)
21. F. Pazos & A. Valencia: Similarity of phylogenetic trees as indicator of protein-protein interaction. *Protein Eng* 14, 609-614 (2001)
22. K. J. Fryxell: The coevolution of gene family trees. *Trends Genet* 12, 364-369 (1996)

23. Curwen V, Eyras E, Andrews T. D, Clarke L, Mongin E, S. M. Searle & M. Clamp: The Ensembl automatic gene annotation system. *Genome Res* 14, 942-950 (2004)
24. Yeh R. F, L. P. Lim & C. B. Burge: Computational inference of homologous gene structures in the human genome. *Genome Res* 11, 803-816 (2001)
25. Thompson J. D, D. G. Higgins & T. J. Gibson: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22, 4673-4680 (1994)
26. J. Felsenstein: PHYLIP (Pyylogeny Inference Package), 3.5c edition. Department of Genetics, University of Washington, Seatile (1998)
27. Kumar S, Tamura K, I. B. Jakobsen & M. Nei: MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17, 1244-1245 (2001)
28. Press W. H, Flannery B. P, S. A. Teukolsky & W. T. Vetterling: Numerical Recipes in C, *Cambridge University Press*. Cambridge, UK. (1998)
29. A. A. Butler & D. LeRoith: Tissue-specific versus generalized gene targeting of the igf1 and igf1r genes and their roles in insulin-like growth factor physiology. *Endocrinology* 142, 1685-1688 (2001)
30. Morison I. M, C. J. Paton & S. D. Cleverley: The imprinted gene and parent-of-origin effect database. *Nucleic Acids Res* 29, 275-276 (2001)
31. B. Tycko & I. M. Morison: Physiological Functions of Imprinted Genes. *J Cell Physiol* 192, 245-258 (2002)
32. Nolan C. M, Killian J. K, J. N. Petitte & R. L. Jirtle: Imprint status of M6P/IGF2R and IGF2 in chickens. *Dev Genes Evol* 211, 179-183 (2001)
33. K. B. Clairmont & M. P. Czech: Chicken and Xenopus mannose 6-phosphate receptors fail to bind insulin-like growth factor II. *J Biol Chem* 264, 16390-16392 (1989)
34. Kelley K. M, Schmidt K. E, Berg L, Sak K, Galima M. M, Gillespie C, Balogh L, Hawayek A, J. A. Reyes & M. Jamison: Comparative endocrinology of the insulin-like growth factor-binding protein. *J Endocrinol* 175, 3-18 (2002)
35. D. M. Irwin: A second insulin gene in fish genomes. *Gen Comp Endocrinol* 135, 150-158 (2004)
36. Leroux L, Desbois P, Lamotte L, Duvillie B, Cordonnier N, Jackerott M, Jami J, D. Bucchini & R. L. Joshi: Compensatory responses in mice carrying a null mutation for Ins1 or Ins2. *Diabetes* 50 S1, S150-53 (2001)
37. Soares M. B, Schon E, Henderson A, Karathanasis S. K, Cate R, Zeitlin S, J. Chirgwin & A. Efstratiadis: RNA-mediated gene duplication: the rat preproinsulin I gene is a functional retroposon. *Mol Cell Biol* 5, 2090-2103 (1985)
38. Newfeld S. J, R. G. Wisotzkey & S. Kumar: Molecular evolution of a developmental pathway: phylogenetic analyses of transforming growth factor-beta family ligands, receptors and Smad signal transducers. *Genetics* 152, 783-795 (1999)
39. N. R. Farid & M. W. Szkudlinski: Minireview: structural and functional evolution of the thyrotropin receptor. *Endocrinology* 145, 4048-4057 (2004)
40. Moyle W. R, Campbell R. K, Myers R. V, Bernard M. P, Y. Han & X. Wang: Co-evolution of ligand-receptor pairs *Nature* 368, 251-255 (1994)
41. van Kesteren R. E, Tensen C. P, Smit A. B, van Minnen J, Kolakowski L. F, Meyerhof W, Richter D, van Heerikhuizen H, E. Vreugdenhil & W. P. Geraerts: Co-evolution of ligand-receptor pairs in the vasopressin/oxytocin superfamily of bioactive peptides. *J Biol Chem* 271, 3619-3626 (1996)

Footnotes: Tonghai Dou and Chaoneng Ji contributed equally to this paper.

Abbreviations: IGF, insulin growth factor; IRS, insulin receptor substrate

Key Words: Co-Evolutionary Relationship, Hormone, Insulin, IGF-I Signaling Pathway, Distance Matrix, Correlation Coefficient

Send correspondence to: Dr Yumin Mao, Institute of Genetics, School of Life Science, Fudan University, Shanghai, PR China, Tel: 86-021-65643723, Fax: 86-021-65642502, E-mail: ymmao@fudan.edu.cn

http://www.bioscience.org/current/vol11.htm