J Mol Evol (2002) 54:725–733
DOI: 10.1007/s00239-001-0072-3

JOURNAL OF MOLECULAR EVOLUTION

Evolutionary Analysis for Functional Divergence of Jak Protein Kinase Domains and Tissue-Specific Genes

Jianying Gu,[1,3] Yufeng Wang,[1,3] Xun Gu[1–3]

[1] Department of Zoology and Genetics, Iowa State University, Ames, IA 50011, USA
[2] Center for Bioinformatics and Biological Statistics, Iowa State University, Ames, IA 50011, USA
[3] Program of Bioinformatics and Computational Biology (BCB), Iowa State University, Ames, IA 50011, USA

Received: 23 August 2001 / Accepted: 8 November 2001

Abstract. Jak (Janus kinase) is a nonreceptor tyrosine kinase, which plays important roles in signal transduction pathways. The unique feature of Jak is that, in addition to a fully functional tyrosine kinase domain (JH1), Jak possesses a pseudokinase domain (JH2). Although JH2 lost its catalytic function, experimental evidence has shown that this domain may have acquired some new but unknown functions. This apparent functional divergence after the (internal) domain duplication may result in dramatic changes of selective constraints at some sites. We conducted a data analysis to test this hypothesis. Our result shows that shifted selective constraints (or shifted evolutionary rates) between the JH1 and the JH2 domains are statistically significant. Predicted amino acid sites by posterior analysis can be classified into two groups: very conserved in JH1 but highly variable in JH2, and *vice versa.* Moreover, we have studied the evolutionary pattern of four tissue-specific genes, Jak1, Jak2, Jak3, and Tyk2, which were generated in the early stages of vertebrates. We found that after the (first) gene duplication, site-specific rate shifts between Jak2/Jak3 and Jak1/Tyk are significant, presumably as a consequence of functional divergence among these genes. The implication of our study for functional genomics is discussed.

Key words: Domain duplication (shuffling) — Gene duplication — Jak — Protein kinase — Functional divergence

Introduction

Jak protein, a group of nonreceptor tyrosine kinases, plays a crucial role in cytokine signaling (Darnell et al. 1994; Leonard and O'Shea 1998). The Jak family consists of four mammalian members: Jak1, Jak2, Jak3, and Tyk2 (Wilks et al. 1991; Harpur et al. 1992; Rane and Reddy 1994). Homologous Jaks from aves and teleosts, as well as a *Drosophila* Jak homologue, *hopscotch,* have also been sequenced (Binari and Perrimon 1994; Conway et al. 1997). The distinguishing feature of the Jak protein family is the existence of tandem kinase (JH1) and pseudokinase (JH2) domains. The tandem kinase domain has all the functional features of a typical tyrosine kinase domain, while the pseudokinase domain has lost the catalytic activity (Guschin et al. 1995; Feng et al. 1997; Zhou et al. 1997).

Among the metazoan protein tyrosine kinases (PTKs), only Jak protein maintains a pseudokinase domain (JH2). The reason for this phenomenon remains unclear. JH2 has all the subdomains that are shared by typical tyrosine kinase domains, indicating a common origin with other kinase domains, e.g., via internal domain duplication. Thus, these two domains in Jak proteins provide an interesting case study of functional divergence in proteins from an evolutionary perspective. It is believed that the

Correspondence to: Xun Gu; *email:* xgu@iastate.edu

lack of some typical motifs, shared by functional kinase domains, in the pseudokinase domain rendered this domain catalytically inactive (Wilks et al. 1991; Frank et al. 1995). On the other hand, knockout evidence has shown either abrogation or stimulation of kinase activity when the JH2 domain is deleted in Jak2, Jak3, or Tyk2 (Velazquez et al. 1995; Candotti et al. 1997; Luo et al. 1997). These facts imply that after the internal domain duplication, the pseudokinase domain has undergone loss of function and gain of function simultaneously. An interesting question in molecular evolution is whether functional alterations in the JH2 domain result in considerable changes in selective constraints (i.e., different evolutionary rate) at those sites involved. As a consequence, some well-conserved sites in JH1 domain become highly variable in the JH2 domain, and vice versa. Such a pattern has been described as "covarion behavior" (Gaucher et al. 2001) or Type I functional divergence (Gu 1999).

In this paper, we use Gu's (1999) method to test the hypothesis that altered selective constraints at some positions occurred between the JH1 and the JH2 domains. Then we study the evolutionary pattern among vertebrate Jak tissue-specific genes and compare this to the JH1/JH2 domain evolution. The implication of our results for functional divergence and specification in Jak/Stat pathway is discussed.

Methods

Data Sets

We searched the Pfam database (http://pfam.wustl.edu/) for sequences that have tandem kinase or pseudokinase domains of Jak genes. For comparison, the kinase domain sequences of the FGFR (fibroblast growth factor receptor) and EGFR (epidermal growth factor receptor) gene families were also downloaded.

To find all available sequences that belong to the Jak gene family, Gapped BLAST and PSI-BLAST searches were performed in several protein databases using the human Jak1 gene as a query sequence (Altschul et al. 1997). After an exhaustive search, partial sequences and redundant sequences were removed from further analysis. The final data set includes 22 complete vertebrate Jak sequences and 1 *Drosophila* homologous gene, *Hopscotch* (see Fig. 4 legend for accession number and species).

Multiple Alignment and Phylogenetic Analysis

The multiple hidden Markov models (HMM) alignment of tandem kinase and pseudokinase domains was obtained from Pfam, followed by manual editing. The multiple alignment of complete Jak sequences was obtained using the program Clustal X (Thompson et al. 1997). These multiple alignments are available upon request. A phylogenetic tree was inferred by the neighbor-joining method (Saitou and Nei 1987) using the software MEGA2.0 (http://www.megasoftware.net/) (Kumar et al. 1994). Parsimony (PAUP) and likelihood (PHYLIP) methods were also used for phylogenetic reconstruction to examine the sensitivity to the tree-making method.

The ratio of nonsynonymous rate to synonymous rate (d_n/d_s) was calculated using a modified version of Gojobori and Nei's method (in the software MEGA2.0). When assuming that the synonymous substitution is virtually neutral, $d_n/d_s > 1$ indicates positive selection, $d_n/d_s < 1$ indicates negative selection, and $d_n/d_s \approx 1$ indicates neutral evolution.

Testing Type I Functional Divergence (Altered Selective Constraint)

Type I functional divergence refers to the evolutionary process that results in altered selective constraints (different evolutionary rates) between two duplicate genes, regardless of the underlying evolutionary mechanisms (Gu 1999, 2001). Its functional–structural basis was well illustrated in the case study of the Caspase family (Wang and Gu 2001). Consider two gene clusters generated by duplication. Briefly speaking, in each cluster an amino acid site is called an F_1 site (functional divergence related) if its evolutionary rate differs from that in the ancestral gene; otherwise it is called an F_0 site (i.e., functional divergence unrelated). Consequently, in the case of two gene clusters, a site can be in either of two states: (1) S_0 or (2) S_1, where S_0 is defined as a site being F_0 in both clusters and S_1 is defined as a site being F_1 in at least one cluster. The *coefficient of Type I functional divergence,* $\theta_{A,B}$, between two gene clusters, A and B, is defined as the probability of a site being in the S_1 state. Thus, a null hypothesis of $\theta = 0$ means that the evolutionary rate is virtually the same between duplicate genes at each site. Note that many models for rate variation among sites are the special case of $\theta = 0$. A probabilistic model has been developed (Gu 1999) to estimate θ by establishing the relationship between this type of functional divergence and the observed amino acid configurations.

Rejection of the null hypothesis (i.e., $\theta > 0$) provides statistical evidence for altered selective constraints of amino acid sites after gene duplication. Thus, we can use a site-specific profile to identify responsible amino acid sites. Specifically we define Q_k to be the posterior probability that site k is in state S_1 ($0 \leq Q_k \leq 1$). A large Q_k indicates a high possibility that the functional constraint (or the evolutionary rate) of a site is different between two clusters.

The software DIVERGE (http://xgul.zool.iastate.edu) was used for the functional divergence analysis.

Functional Distance Analysis

One limitation of the two-cluster analysis described above is that it cannot test whether one duplicate gene has more shifted evolutionary rates than the other one. This issue is addressed by a simple method described below that can be applied when more than two homologous gene clusters are available (Wang and Gu 2001). First, the Type I functional distance between any two clusters is defined as $d_F = -\ln(1 - \theta)$. Under the assumption of independence, Wang and Gu (2001) have shown that d_F is additive, i.e., for two clusters, A and B, $d_F(A,B) = b_F(A) + b_F(B)$, where $b_F(x)$ is the functional branch length of a given gene cluster x. A large b_F value for a gene cluster indicates that the evolutionary conservation may be shifted at many sites.

The estimated coefficients of Type I functional divergence (θ) for all the pairs of clusters can be used to create a matrix of d_F values. Given this matrix, a standard least-squares method can be implemented based on the formula $d_F(A,B) = b_F(A) + b_F(B)$ to estimate b_F for each gene cluster. If $b_F \approx 0$, it indicates that the evolutionary rate of each site in this duplicate gene has remained nearly the same since the gene duplication event, indicative that the derived state is more similar to the ancestral state for this particular cluster.

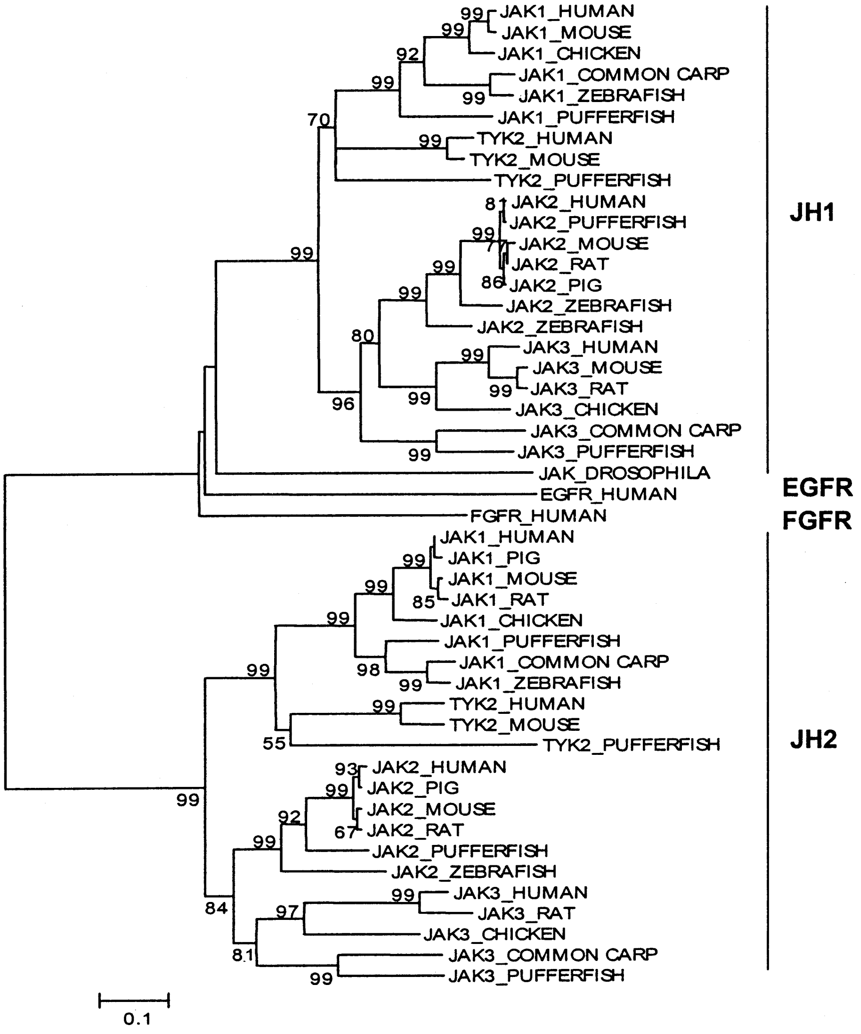

Fig. 1. The NJ tree of Jaks, FGFRs, and EGFRs based on the sequence alignment of kinase domains. The statistical reliability of the inferred tree topology was assessed by the bootstrap technique (Felsenstein 1985).

Results

Pattern of Functional Divergence Between Tandem Kinase and Pseudokinase Domains

The Kinase and Pseudokinase Domains in Jaks

In addition to the regular kinase domain (JH1), Jak proteins contain a pseudokinase domain (JH2) with a functional role that has not been clearly determined (Aringer et al. 1999). To explore the evolutionary pattern of the JH1 and JH2 domains, we reconstructed a neighbor-joining (NJ) tree, including Jaks and two closely related protein tyrosine kinases, FGFR and EGFR. The inferred phylogeny shows that the tandem kinase (JH1) and pseudokinase (JH2) domains are evolutionarily distinct (Fig. 1). Indeed, the tandem kinase domains (JH1) in Jaks appear to be more closely related to the functional kinase domains of FGFRs and EGFRs, while the pseudokinase domains (JH2) of Jaks form a unique clade. In fact, we have found that the JH2 domain was generated before the emergence of most member genes of the Protein tyrosine kinase supergene family (data not shown).

It has been suggested that the pseudokinase domains (JH2) no longer exhibit the catalytic activity but may have acquired some new functions (Aringer et al. 1999). Thus, it is interesting to test whether this functional divergence resulted in shifted selective constraints (differ-

Table 1. θ_{AB} values for all combinations of pairwise comparisons of Jaks, FGFRs, and EGFRs[a]

	JH1	JH2	FGFR	EGFR
JH1		0.412 ± 0.049	0.222 ± 0.055	0.496 ± 0.071
JH2	0.531		0.566 ± 0.080	0.561 ± 0.080
FGFR	0.251	0.835		0.432 ± 0.096
EGFR	0.685	0.823	0.566	

[a] JH1, the tandem kinase domain in Jak; JH2, the pseudokinase domain in Jak. The upper diagonal shows the θ_{AB} values for all pairwise comparisons of Jaks, FGFRs, and EGFRs, where θ_{AB} is defined as the coefficient of the functional divergence between cluster A and cluster B. The lower diagonal shows $d_F(A,B) = -\ln(1 - \theta)$, which is defined as the functional distance between cluster A and cluster B.

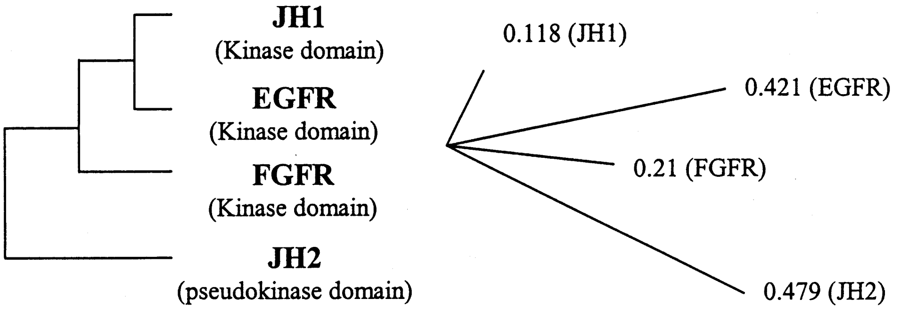

Fig. 2. The tree-like topology of kinase domains of Jaks, FGFRs, and EGFRs in terms of the functional distance b_F.

ent evolutionary rates) at some sites between the tandem kinase (JH1) and the pseudokinase (JH2) domains. To this end, we estimated the coefficient of functional divergence (θ) between the JH1, JH2, EGFR, and FGFR domains, as shown in the upper diagonal in Table 1. All of the estimates are significantly greater than 0 ($p < 0.01$). In particular, $\theta_{JH1,JH2} \pm SE = 0.412 \pm 0.049$, providing statistical evidence for supporting the hypothesis of altered selective constraints between the tandem kinase (JH1) and pseudokinase (JH2) domains in Jak proteins.

To test further the hypothesis that JH2 is more functionally divergent than JH1, we conducted a functional distance analysis (Wang and Gu 2001). The functional distances between domains [$d_F = -\ln(1 - \theta)$] are shown in the lower diagonal in Table 1. Subsequently, the functional branch length (b_F) for each domain was estimated by the least-squares method, which can be illustrated as a star-like tree (Fig. 2). The null hypothesis of equal functional branch lengths is rejected ($p < 0.05$). Interestingly, the functional branch length of the pseudokinase domain (JH2) of Jak proteins is about four times greater than that of the kinase domain (JH1), indicating that the distinct functional role of JH2 is likely generated by the episodic evolution of kinase domains.

Important Amino Acid Sites for Altered Functional Constraints Between JH1 and JH2

A site-specific profile based on the posterior probability (Q_k) is used to identify critical amino acid sites that are responsible for functional divergence between the tandem kinase (JH1) and the pseudokinase (JH2) domains. Among 212 amino acid sites, 154 (73%) sites have $Q_k < 0.5$, implying a low probability of contribution to functional divergence. For the remaining 58 amino acid sites, with $Q_k > 0.5$, 21 of them show a very high probability of being functional divergence related ($Q_k > 0.9$). These 21 sites can be definitively grouped into two categories: (I) conserved in the tandem kinase (JH1) domain but variable in the pseudokinase (JH2) domain and (II) conserved in the pseudokinase domain but variable in the tandem kinase domain (Table 2).

Table 2. Predicted critical amino acid sites responsible for functional divergence between the tandem kinase and the pseudokinase domain in Jaks[a]

Position (k)	Q_k
I	
26	0.997
203	0.992
30	0.975
20	0.968
46	0.967
23	0.966
162	0.963
70	0.959
137	**0.957**
21	0.922
159	0.913
196	0.906
II	
38	0.998
47	0.967
41	0.960
190	0.959
105	0.956
103	**0.954**
44	0.931
121	0.918
87	0.906

[a] Category I, conserved in the tandem kinase domain and variable in the pseudokinase domain; category II, conserved in the pseudokinase domain and variable in the tandem kinase domain.

Category I. Of the 12 sites belonging to this category, 1 site, site 137 ($Q_{137} = 0.957$), has been demonstrated to be a determining site for the function of the tandem kinase domain (JH1), corresponding to the second tyrosine (highlighted **Y**) of a conserved (E/D)Y**Y** motif in the Jak2 protein. This motif, which is located in the activation loop of Jak2, regulates the kinase activity by autophosphorylation (Feng et al. 1997). In Tyk2, these two consecutive tyrosines (Y**Y**), have also been identified as phosphorylation sites (Gauzzi et al. 1996).

Interestingly, the multiple alignments clearly show that site 137 is invariant in the tandem kinase (JH1) domains. In contrast, the same position in the pseudokinase domains (JH2) has a variety of amino acids with very different chemical properties. For example, some JH2 domains have amino acids with nonpolar side chains such as glycine, alanine, and proline, and some of them have uncharged polar amino acids such as serine and threonine (Fig. 3A). This observation can be ex-

	(A)			(B)	
		11112			1111
	position (*k*)	222234735690		position (*k*)	344480029
		013606079263			814773510
JH1	HUMAN	DPDGALPYYSRC	Jak1	HUMAN	KERYQGRST
	MOUSE	DPDGALPYYSRC		MOUSE	KERYQGRST
	CHICKEN	DPDGALPYYSRC		CHICKEN	KERYQGRSV
	CARP	DPDGALPYYSRC		CARP	WHRYTARNV
	ZEBRAFISH	DPDGALPYYSRC		ZEBRAFISH	WHRYTGRNV
	PUFFERFISH	DPDGALPYYSRC		PUFFERFISH	SDKFTGKNL
	HUMAN	DPDGALPYSSRC	Jak2	HUMAN	EEKQKGKNL
	MOUSE	DPDGALPYSSRC		MOUSE	EEKQKGKNL
	RAT	DPDGALPYSSRC		RAT	EEKQKGKNL
	PIG	DPDGALPYSSRC		PIG	EEKQKGKNL
	PUFFERFISH	DPDGALPYSSRC		PUFFERFISH	EEKQKGKNL
	ZEBRAFISH	DPDGALPYSSRC		ZEBRAFISH	EEKQKAKSL
	HUMAN	DPHGALPYSSRC	Jak3	HUMAN	QQKHRGRSL
	MOUSE	DPDGALPYSSRC		MOUSE	QQKHRGRSL
	RAT	DPDGALPYSSRC		RAT	QQKHRGRSL
	CHICKEN	DPDGALPYSSRC		CHICKEN	EQHQTGQSL
	CARP	DPDGALPYSSRC		CARP	QQSHRQMSF
	PUFFERFISH	DPDGALPYSSRC		PUFFERFISH	KKSHRQLSI
	HUMAN	DPDGALPYYSRC	Tyk2	HUMAN	KDRYQHHNL
	MOUSE	DPDGALPYYSRC		MOUSE	QERYQHQNL
	PUFFERFISH	DPDGALPYSSRC		PUFFERFISH	INKDQHKRL
JH2	HUMAN	MDEKIVESSARK	Jak1	HUMAN	FAMSWEKRF
	PIG	MDEKIVESSARK		PIG	FAMSWEKRF
	MOUSE	LDEKIVETSARK		MOUSE	FAMSWEKRF
	RAT	LDEKIVETSARK		RAT	FAMSWEKRF
	CHICKEN	LNNELVESSAMK		CHICKEN	FAMSWEKRF
	CARP	KLYEIIQSSAQE		CARP	FAMSWEKRF
	ZEBRAFISH	KPYEVIQSTAQD		ZEBRAFISH	FAMSWEKRF
	PUFFERFISH	RVSEVVQTSAQT		PUFFERFISH	FAMSWEKRF
	HUMAN	REQELLKPNTQA	Jak2	HUMAN	FAMSWENRF
	MOUSE	REQKLLKPNTQT		MOUSE	FAMSWEKRF
	RAT	REQELLKPTTQT		RAT	FAMSWEKRF
	PIG	REQELLKPNTQT		PIG	FAMSWEKRF
	PUFFERFISH	KELQVLKPSAQI		PUFFERFISH	FAMSWEKRF
	ZEBRAFISH	REEKVLRPSAQT		ZEBRAFISH	FAMSWEKRF
	HUMAN	HEEKLVHSSAQT	Jak3	HUMAN	FAMSWEKRF
	RAT	REEDLVYSNAQT		RAT	FAMSWEKRF
	CHICKEN	RDEQVLRAAAQS		CHICKEN	FAMSWEKRF
	CARP	TDVTLIKGECNT		CARP	FAMSWEKRF
	PUFFERFISH	SNGRFFEGTSQT		PUFFERFISH	FAMSWENRF
	HUMAN	RVREVVESSMRP	Tyk2	HUMAN	FAMSWEKRF
	MOUSE	RVSQVVESGTQP		MOUSE	FAMSWEKRF
	PUFFERFISH	QVSDVLKTRPRK		PUFFERFISH	FVMSWEKRF

Fig. 3. Functional divergence related amino acid sites candidate. (A) Category I: conserved in tandem kinase domains (JH1) and variable in pseudokinase domains (JH2). (B) Category II: conserved in pseudokinase domains (JH2) and variable in tandem kinase domains (JH1).

plained as a relaxed selective constraint that was caused by loss of function in phosphorylation in the JH2 domains.

Category II. Nine predicted sites belong to this category (Fig. 3B). Among them, site 103 is predicted to be highly functional divergence related ($Q_{103} = 0.954$). Experimental data show that a glutamic acid (E)-to-lysine (K) replacement occurring at this site in the pseudokinase (JH2) domain hyperactivated the Jak–Stat pathway in *Drosophila* and mammalian species (Luo et al. 1997). It seems likely that after the internal domain duplication, the tandem kinase domain (JH1) largely maintained the original catalytic function, while the pseudokinase domain (JH2) may have achieved some unidentified new functions, resulting in a set of JH2-specific conserved sites.

Pattern of Functional Divergence Between Jak Member Genes

Functional Divergence of the Jak Gene Family

Figure 4 shows the neighbor-joining (NJ) tree of four Jak gene members (Jak1, Jak2, Jak3, and Tyk2). The parsimony method (PAUP) and likelihood method (PHYLIP) give essentially the same topology (data not shown). The phylogenetic analysis indicates that Jak member genes were generated by two gene duplications in the early stage of vertebrates, i.e., before the emergence of teleosts. The first gene duplication resulted in the common ancestor of Jak2/3 and Tyk2/Jak1, and the second one resulted in the current four member genes. Note that Wang and Gu (2000) have shown that many tissue-specific gene families show a similar pattern, rais-

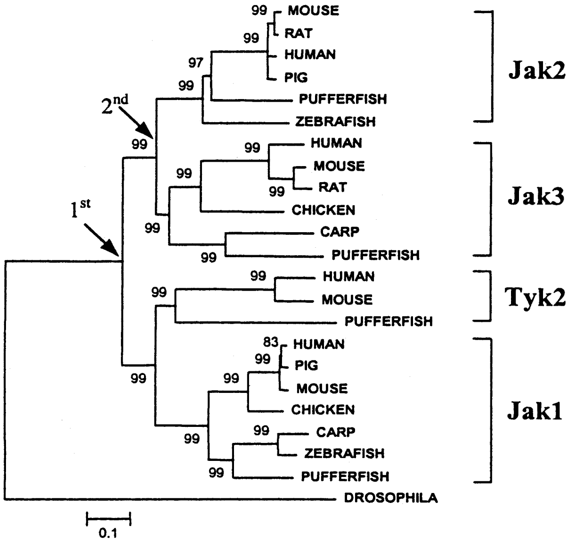

Fig. 4. The phylogenetic tree of the Jak gene family. The neighbor-joining algorithm was used to infer the topology based on the multiple sequence alignment with Poisson distance. Bootstrap scores >50% are presented. 1st and 2nd represent the time points of two rounds of gene duplications. Accession numbers: Jak2—L16956 (mouse), U13396 (rat), AF005216 (human), AB006011 (pig), AF090382 (pufferfish), AJ005690 (zebrafish); Jak3—U09607 (human), L40172 (mouse), D28508 (rat), AF034576 (chicken), AF148993 (carp), AF091238 (pufferfish); Tyk2—X54637 (human), AF173032 (mouse), AF090383 (pufferfish); Jak1—M64174 (human), AB036335 (pig), M33425 (mouse), AF096264 (chicken), L24895 (carp), U82980 (zebrafish), U53213 (pufferfish); *hopscotch*—L26975 *(Drosophila).*

ing the possibility of a large-scale duplication(s) in early vertebrates.

As nucleotide change in the regulatory region after gene duplication may be the first step in duplicate gene preservation (Force et al. 1999), amino acid replacements are responsible for the functional divergence for specification among tissue-specific isoforms. To explore whether (site-specific) altered selective constraints occurred during Jak family evolution, we estimated the θ value between Jak member genes, as well as between two groups (Jak2/3 and Tyk2/Jak1), that was generated by the first-round gene duplication. The estimation is based on the phylogenetic tree in Fig. 4. We found that (1) (site-specific) altered selective constraint after the first gene duplication is statistically significant, and (2) the θ value between Jak2 and Jak3 is 0.019 ± 0.059, indicating, overall, no significant site-specific shift of evolutionary rate between them (Table 3A).

Furthermore, we estimated the coefficients of Type I functional divergence between Jak genes for three separate regions: (1) the tandem kinase domain (JH1), (2) the pseudokinase domain (JH2), and (3) the surrounding region excluding the JH1 and JH2 domains (Table 3B). Our results can be summarized as follows. First, after first-round gene duplication, the site-specific altered selective constraint in the JH1 domain is statistically significant ($p < 0.01$) but not significant in the JH2 domain; there is weak evidence in the surrounding region ($p \sim 0.05$). Second, after the second gene duplication, no statistical evidence is observed for altered selective constraint in any domain. Since the estimate of θ in the JH1 domain tends to be higher than those for the JH2 domain or the surrounding region, we conclude that the JH1 domain is mainly responsible for functional divergence among Jak tissue-specific genes.

Isoform-Specific Functional Divergence of the Jak Gene Family

The functional branch lengths (b_F) of Jak1, Jak2, and Jak3 can be estimated from the pairwise estimates of θ between Jak1, Jak2, and Jak3, respectively. Table 4 shows the results for the whole length of Jak proteins, and three domains (JH1, JH2, and the surrounding region), respectively. Interestingly, the level of altered selective constraints of member genes, measured by b_F, follows b_F (Jak1) > b_F (Jak2) > b_F (Jak3), while the level of altered selective constraints of domains follows JH1 > surrounding region > JH2. In particular, the b_F for Jak3 is virtually zero for all three domains, whereas JH2 (the pseudokinase domain) shows no significant functional branch length in Jak2 and Jak3.

Table 4 also includes the ratios of nonsynonymous-to-synonymous rates (d_n/d_s) of Jak member genes, based on human–mouse orthologous members of a gene, which can be used to measure the difference of selective constraints among members of a gene family (Tsunoyama et al. 1998). Interestingly, Jak3, which has virtually $b_F = 0$, shows the highest d_n/d_s ratio. This observed negative association between d_n/d_s and b_F can be interpreted as follows. In the early stage after gene duplication, the functional divergence may occur in one of two lineages (measured by a high b_F value). If this process leads to the acquisition of some new functions, a stronger functional constraint (measured by a low d_n/d_s value) is expected.

In summary, during Jak tissue-specific gene evolution, Type I functional divergence (i.e., site-specific altered selective constraints) occurs mainly in the JH1 domain of the Jak1 and Jak2 genes. As the Jak1 and Jak2 genes may have evolved specialized functions in their kinase domains, Jak3 is more likely to have inherited the ancestral function.

Predicting Important Amino Acid Sites for Type I Functional Divergence

Critical amino acid sites responsible for the Type I functional divergence among Jak member genes can be predicted by Q_k, the posterior probability of being functional divergence related at site k (Gu 1999). Figure 5 shows the site-specific profile between the Jak1/Tyk2 and the Jak2/Jak3 clusters, measured by Q_k. It clearly shows that, after the first gene duplication, only a small portion of sites has undergone shifted rates, as indicated by high posterior probabilities. These sites may be considered as a prime candidates for further functional assays.

Table 3. Pairwise coefficients of functional divergence (θ_{AB}) for members of the Jak gene family: (A) full-length Jak proteins; (B) different regions of Jak proteins

A[a]

Cluster 1	Cluster 2	Gene duplication[b]	θ ± SE (ML)	LRT[c]	*p* value
Jak1	Jak2	I	0.208 ± 0.056	13.609	<0.0005
Jak1	Jak3	I	0.158 ± 0.051	9.502	0.002
Jak2	Jak3	II	0.019 ± 0.059	0.110	0.75
Jak1/Tyk2	Jak2/Jak3	I	0.072 ± 0.027	7.292	0.007
Jak1	Jak2/Jak3	I	0.147 ± 0.039	14.367	<0.0005

B

			θ ± SE			
Cluster 1	Cluster 2	Gene duplication	JH1	JH2	Surrounding	Full-length
Jak1	Jak2	I	0.435 ± 0.140	0.214 ± 0.125	0.294 ± 0.100	0.208 ± 0.056
Jak1	Jak3	I	0.231 ± 0.086	0.052 ± 0.143	0.178 ± 0.089	0.158 ± 0.051
Jak2	Jak3	II	0.217 ± 0.150	0.001 ± 0.022	0.022 ± 0.100	0.019 ± 0.059
Jak1/Tyk2	Jak2/Jak3	I	0.122 ± 0.050	0.094 ± 0.068	0.101 ± 0.053	0.072 ± 0.027
Jak1	Jak2/Jak3	I	0.174 ± 0.067	0.086 ± 0.097	0.225 ± 0.074	0.147 ± 0.039

[a] As a single gene cluster, Tyk2 was excluded from the analysis due to the insufficient number of sequences.

[b] I represents the first gene duplication; II, the second gene duplication.

[c] Likelihood ratio statistic.

Table 4. Functional branch lengths (b_F) of different regions in three Jak isoforms

	JH1	JH2	Surrounding	Full length	d_n/d_s (full-length)
Jak1	0.295	0.147	0.261	0.193	0.074
Jak2	0.277	0.095	0.087	0.040	0.089
Jak3	−0.032	−0.094	−0.065	−0.021	0.274

Discussion

Domain and gene duplications are both important for Jak gene family proliferation and evolution. After gene (or domain) duplication, one copy mainly retains the original function, whereas the other copy is under relaxed evolutionary constraint that may result in functional divergence and/or specification (Ohno 1970; Li 1983). In this paper, we have investigated the evolutionary pattern of the tandem kinase (JH1) and pseudokinase (JH2) domains, a unique feature of the supergene family of tyrosine protein kinases. Our results can be summarized as follows.

In the very early stage of animals, an internal duplication had occurred within the ancestor of the Jak gene, producing two kinase domains. As one domain (JH1) largely retains the original kinase activity, the other (JH2) is free to accumulate amino acid replacements because of functional redundancy. When some key sites were replaced by other amino acids without serious deleterious effects on the survival, the kinase function of the JH2 domain was lost. Consequently, some well-conversed sites in functional kinase domains turn out to be highly variable. Meanwhile, the JH2 domain appears to have acquired some new functions that may explain its long-term existence during evolution. Despite the lack of substantial evidence for a physiological role of JH2, we indeed observed a group of sites in the JH2 domain to be very conserved, whereas they were variable in the JH1 domain.

It seems likely that the functional divergence between the JH1 and the JH2 domains was complete before the origin of vertebrates. In the early vertebrate lineage, two rounds of gene duplications generated four tissue-specific vertebrate isoforms. Type I functional divergence (altered selective constraint) was significant after the first round of gene duplication. This divergence occurred mainly in the JH1 domain, and not in the JH2 domain, indicating that the JH1 domain probably plays a major role in functional specification among tissue-specific isoforms.

Since our analysis is based on the amino acid alignment, reliability of the multiple alignment is crucial for our interpretation. It has been indicated that the coefficient of Type I functional divergence (θ) can be overestimated by the misaligned sequences (Gu 1999, 2001). Indeed, we found a larger estimate of θ (θ = 0.615) between JH1 and JH2 for the unedited Pfam domain HMM alignment, which includes several apparent misaligned sites. After making manual corrections at these sites according to the consensus sequence of protein kinase domains that have been confirmed by the secondary structure, the estimate of θ was reduced to 0.439. Furthermore, we tested the effect of the alignment by shift-

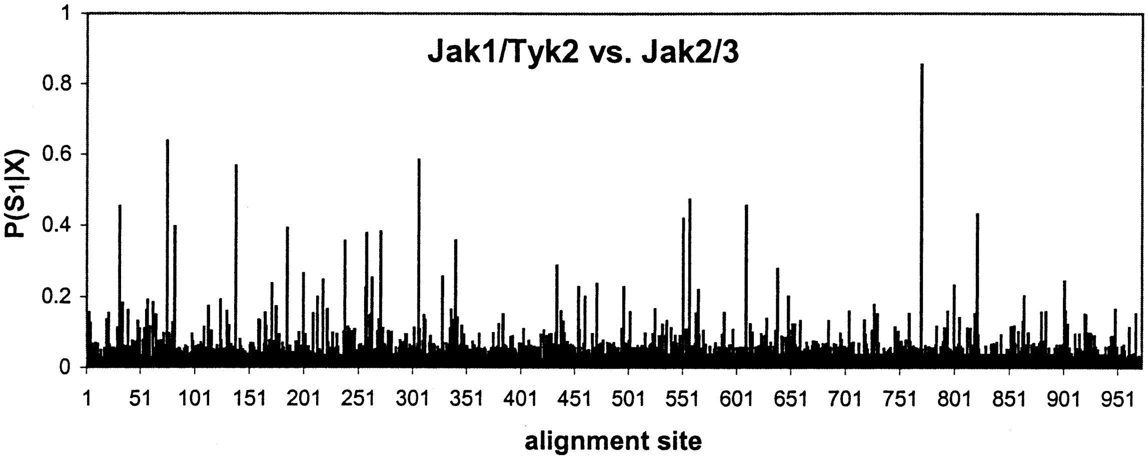

Fig. 5. Site-specific profiles (Q_k) for clusters Jak1/Tyk2 vs Jak2/Jak3.

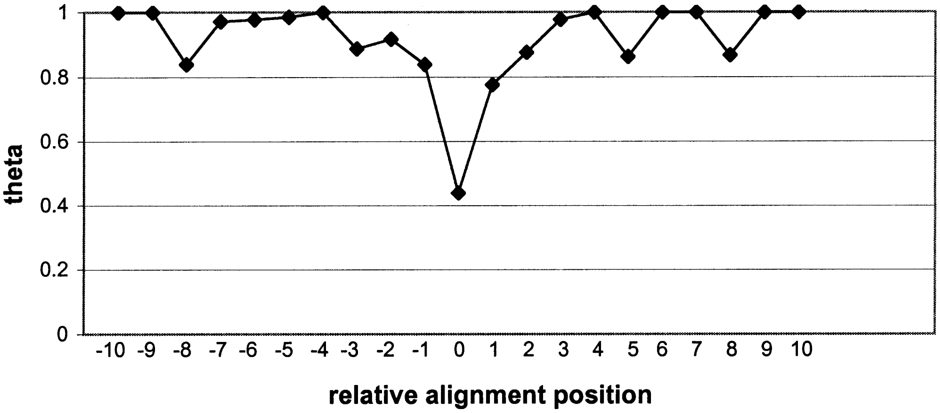

Fig. 6. Correlation of the θ value with the sequence alignment. The *y* axis represents different θ values corresponding to the alignment by sliding positions (−10 to +10) relative to the manually adjusted alignment of kinase domains in Jaks.

ing the alignment window between JH1 and JH2 from −10 to +10 positions. Figure 6 shows that all estimates of θ that resulted from the shifted alignment are extremely high. Our study suggests that a reasonably good alignment is a prerequisite for estimating the level of functional divergence.

Gu's (1999) method applies only to the Type I functional divergence that resulted in site-specific altered selective constraints among homologous genes within a gene family. For example, the amino acid site 137 shows a typical Type I pattern, i.e., invariant tyrosine in all tandem kinase domains but highly variable in pseudokinase domains. It should be noted that this is only one of many approaches to detect functional divergence from the evolutionary perspective. As indicated by many authors (e.g., Casari et al. 1995; Livingstone et al. 1996; Gaucher et al. 2001), functional divergence may result in dramatic changes in amino acid properties but not in selective constraints, which is also called Type II functional divergence (Gu 1999). For instance, cluster-specific sites (or diagnosis sites) usually refer to sites that, though highly conserved within gene clusters, consist of amino acids that are dramatically different (e.g., positively charged vs negatively charged) between clusters. Another well-cited example is that site-by-site dependence may be related to the site interaction in the protein's three-dimensional structure (Pollock et al. 1999). A sophisticated model that includes all these evolutionary aspects of functional divergence of proteins is certainly desirable. However, because of the large number of unknown parameters, its efficiency in practice often becomes the major concern. Therefore, a specific model is useful to test the statistical significance of a specific evolutionary perspective. Combined with previous studies (e.g., Gu 1999; Gaucher et al. 2001; Wang and Gu 2001), our analysis indicates that site-specific altered selective constraint occurs after gene duplications as a result of functional divergence in protein evolution.

Acknowledgments. We are grateful to Dr. William Nordstrom and Peter Vedell for constructive comments. This study was supported by NIH Grant RO1 GM62118 (to X. G). X. Gu is the DuPont Young Professor.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Aringer M, Cheng A, Nelson JW, Chen M, Sudarshan C, Zhou YJ, O'Shea JJ (1999) Janus kinases and their role in growth and disease. Life Sci 64:2173–2186

Binari R, Perrimon N (1994) Stripe-specific regulation of pair-rule genes by hopscotch, a putative Jak family tyrosine kinase in Drosophila. Genes Dev 8:300–312

Candotti F, Oakes SA, Johnston JA, Giliani S, Schumacher RF, Mella P, Fiorini M, Ugazio AG, Badolato R, Notarangelo LD, Bozzi F, Macchi P, Strina D, Vezzoni P, Blaese RM, O'Shea JJ, Villa A (1997) Structural and functional basis for JAK3-deficient severe combined immunodeficiency. Blood 90:3996–4003

Casari G, Sander C, Valencia A (1995) A method to predict functional residues in proteins. Nature Struct Biol 2:171–178

Conway G, Margoliath A, Wong-Madden S, Roberts RJ, Gilbert W (1997) Jak1 kinase is required for cell migrations and anterior specification in zebrafish embryos. Proc Natl Acad Sci USA 94: 3082–3087

Darnell JE Jr, Kerr IM, Stark GR (1994) Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. Science 264:1415–1421

Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783–791

Feng J, Witthuhn BA, Matsuda T, Kohlhuber F, Kerr IM, Ihle JN (1997) Activation of Jak2 catalytic activity requires phosphorylation of Y1007 in the kinase activation loop. Mol Cell Biol 17:2497–2501

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531–1545

Frank SJ, Yi W, Zhao Y, Goldsmith JF, Gilliland G, Jiang J, Sakai I, Kraft AS (1995) Regions of the JAK2 tyrosine kinase required for coupling to the growth hormone receptor. J Biol Chem 270:14776–14785

Gaucher EA, Miyamoto MM, Benner SA (2001) Function-structure analysis of proteins using covarion-based evolutionary approaches: Elongation factors. Proc Natl Acad Sci USA 98:548–552

Gauzzi MC, Velazquez L, McKendry R, Mogensen KE, Fellous M, Pellegrini S (1996) Interferon-α-dependent activation of Tyk2 requires phosphorylation of positive regulatory tyrosines by another kianse. J Biol Chem 271:20494–20500

Gu X (1999) Statistical methods for testing functional divergence after gene duplication. Mol Biol Evol 16:1664–1674

Guschin D, Rogers N, Briscoe J, et al. (1995) A major role for the protein tyrosine kinase JAK1 in the JAK/STAT signal transduction pathway in response to interleukin-6. EMBO J 14:1421–1429

Harpur AG, Andres AC, Ziemiecki A, Aston RR, Wilks AF (1992) Jak2, a third member of the JAK family of protein tyrosine kinases. Oncogene 7:1347–1353

Kumar S, Tamura K, Nei M (1994) MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. Comput Appl Biosci 10:189–191

Landgraf R, Fischer D, Eisenberg D (1999) Analysis of heregulin symmetry by weighted evolutionary tracing. Protein Eng 12:943–951

Leonard WJ, O'Shea JJ (1998) Jaks and STATs: Biological implications. Annu Rev Immunol 16:293–322

Li WH (1983) Evolution of duplicate genes and pseudogenes. In: Nei M, Keohn RK (eds) Evolution of genes and proteins. Sinauer Associates, Sunderland, MA, pp 14–37

Li WH (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J Mol Evol 36:96–99

Lichtarge O, Bourne HR, Cohen FE (1996) An evolutionary trace method defines binding surfaces common to protein families. J Mol Biol 257:342–358

Livingstone CD, Barton GJ (1996) Identification of functional residues and secondary structure from protein multiple sequence alignment. Methods Enzymol 266:497–512

Luo H, Rose P, Barber D, Hanratty WP, Lee S, Roberts TM, D'Andrea AD, Dearolf CR (1997) Mutation in the Jak kinase JH2 domain hyperactivates Drosophila and mammalian Jak-Stat pathways. Mol Cell Biol 17:1562–1571

Ohno S (1970) Evolution by gene duplication. Springer, New York

Pollock DD, Taylor WR, Goldman N (1999) Coevolving protein residues: Maximum likelihood identification and relationship to structure. J Mol Biol 287:187–198

Rane SG, Reddy EP (1994) Jak3: a novel JAK kinase associated with terminal differentiation of hematopoietic cells. Oncogene 9:2415–2423

Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Tsunoyama K, Gojobori T (1998) Evolution of nicotinic acetylcholine receptor subunits. Mol Biol Evol 15:518–527

Velazquez L, Mogensen KE, Barbieri G, Fellous M, Uze G, Pellegrini S (1995) Distinct domains of the protein tyrosine kinase tyk2 required for binding of interferon-alpha/beta and for signal transduction. J Biol Chem 270:3327–3334

Wang Y, Gu X (2001) Functional divergence in Caspase gene family and altered functional constraints: Statistical analysis and prediction. Genetics 158:1311–1320

Wilks AF, Harpur AG, Kurban RR, Ralph SJ, Zurcher G, Ziemiecki A (1991) Two novel protein-tyrosine kinases, each with a second phosphotransferase-related catalytic domain, define a new class of protein kinase. Mol Cell Biol 11:2057–2065

Zhou YJ, Hanson EP, Chen YQ, Magnuson K, Chen M, Swann PG, Wange RL, Changelian PS, O'Shea JJ (1997) Distinct tyrosine phosphorylation sites in Jak3 kinase domain positively and negatively regulate its enzymatic activity. Proc Natl Acad Sci USA 94:13850–13855