



Evolution of the *Fgf* and *Fgfr* gene families

Nobuyuki Itoh¹ and David M. Ornitz²

¹Department of Genetic Biochemistry, Kyoto University Graduate School of Pharmaceutical Sciences, Yoshida-Shimoadachi, Sakyo, Kyoto 606-8501, Japan

²Department of Molecular Biology and Pharmacology, Washington University Medical School, 660 S. Euclid Avenue, St Louis, MO 63110, USA

Fibroblast growth factors (Fgfs) and Fgf receptors (Fgfrs) comprise a signaling system that is conserved throughout metazoan evolution. Twenty-two *Fgfs* and four *Fgfrs* have been identified in humans and mice. During evolution, the *Fgf* family appears to have expanded in two phases. In the first phase, during early metazoan evolution, *Fgfs* expanded from two or three to six genes by gene duplication. In the second phase, during the evolution of early vertebrates, the *Fgf* family expanded by two large-scale gen(om)e duplications. By contrast, the *Fgfr* family has expanded only in the second phase. However, the acquisition of alternative splicing by *Fgfrs* has increased their functional diversity. The mechanisms that regulate alternative splicing have been conserved since the divergences of echinoderms and vertebrates. The expansion of the *Fgf* and *Fgfr* gene families has enabled this signaling system to acquire functional diversity and, therefore, an almost ubiquitous involvement in developmental and physiological processes.

Fibroblast growth factors (Fgfs) are polypeptide growth factors with diverse biological activities. Most Fgfs mediate their biological responses by binding to and activating cell surface tyrosine kinase Fgf receptors (Fgfrs) [1,2]. Each Fgfr binds to and is activated by a unique subset of Fgfs, the specificity of which is further regulated by the alternative splicing of the genes encoding the Fgfrs. These genes and those encoding Fgfs have been identified in multicellular organisms ranging from the nematode, *Caenorhabditis elegans*, to the mouse, *Mus musculus*, and the human, *Homo sapiens*, but have not been identified in unicellular organisms such as *Escherichia coli* and *Saccharomyces cerevisiae*. Two *Fgf* genes and one *Fgfr* gene are found in *C. elegans*, whereas, 22 *Fgf* and four *Fgfr* genes are found in humans and mice, indicating that both the *Fgf* and *Fgfr* gene families have greatly expanded during the evolution of primitive metazoa to vertebrates. By their expansion, the FGF signaling system acquired functional diversity in both developmental and physiological processes. In this article, we review of the evolutionary processes that might have led to the expansion of the *Fgf* and *Fgfr* gene families. Human FGFs,

which comprise ~150–300 amino acids, have a conserved ~120-amino acid residue core, with ~30–60% amino acid identity. The overall amino-acid identity among human FGFRs is 56–71% [2]. Traditionally, a group of genes can be considered a gene family if all members of the group have >50% amino acid identity, and an alignable group of genes is considered a superfamily if members of the group have <50% amino acid identity [3]. However, in this article, we use the term ‘gene family’ to encompass both types of groups.

Fibroblast growth factors

Fgf1 and Fgf2 were originally isolated as mitogens for fibroblasts from the brain and pituitary [1,2]. Fgfs are widely expressed in developing and adult tissues and have various biological activities both *in vivo* and *in vitro*, including roles in angiogenesis, mitogenesis, cellular differentiation, cell migration and tissue-injury repair. Fgfs interact with heparin or heparan sulfate proteoglycans, which stabilize Fgfs and prevent thermal denaturation and proteolysis, and are required for Fgfs to activate Fgfrs effectively [1,4]. Fgf binding proteins increase the bioavailability of Fgfs and might facilitate receptor activation [5].

The *Fgf* gene family in vertebrates

FGF1–FGF23 have been identified in humans and mice. However, *FGF19* is the human ortholog of mouse *Fgf15* [1]. In total, the human–mouse FGF family consists of 22 members [1]. Most FGFs (FGFs 3–8, 10, 17–19, 21 and 23) have N-terminal signal peptides and are readily secreted from cells. By contrast, FGFs 9, 16 and 20 lack obvious, cleavable N-terminal signal peptides but are nevertheless secreted. These FGFs have an uncleavable N-terminal hydrophobic sequence that is required for secretion. FGF1 and FGF2 also lack signal peptides, however, unlike FGFs 9, 16 and 20, are not secreted but can be released from damaged cells or by an exocytotic mechanism that is independent of the endoplasmic reticulum-Golgi pathway [1]. FGF22, with a putative N-terminal signal peptide, remains attached to the cell surface rather than being secreted. FGFs 11–14 lack signal peptides, remain intracellular and function within cells in a receptor-independent manner. These FGFs might be intracellular components of a tissue-specific

Corresponding author: Nobuyuki Itoh (itohnobu@pharm.kyoto-u.ac.jp).

protein-kinase signaling module and seem to share structural, but not functional, homology with other FGFs [6].

Phylogenetic analyses divide the human *FGF* gene family into seven subfamilies (Figure 1). Genome sequences of other vertebrates including the rat, *Rattus norvegicus* [7], puffer fish, *Fugu rubripes* [8] and zebrafish, *Danio rerio* have been almost completed, and orthologs of most of the human *FGFs* have been identified in these genomes (N. Itoh, unpublished). These findings indicate that no *Fgf* gene was acquired in the mammalian lineage after the divergence of mammalian from non-mammalian vertebrates. Comparisons of mammal genes with those of teleost fish have shown that in teleosts, including zebrafish, there are often two homologs of the mammalian equivalent. This suggests that there has been an additional genome duplication shortly before the teleost radiation. This duplication must have been either partial or a whole genome duplication followed by rapid gene loss because gene duplications account for only ~20% of the zebrafish genes examined [9]. Although the total number of *fgf* genes in the zebrafish genome remains to be elucidated, there are at least 26 *fgf* genes in the zebrafish genome (N. Itoh, unpublished).

Chromosomal locations of human *FGF* genes

As described previously, the human *FGF* gene family is divided into seven subfamilies (Figure 1). Members of each subfamily share increased sequence similarity, and biochemical and developmental properties [1,2]. The chromosomal locations of all human *FGF* genes, except for *FGF16*, are known (Table 1). Most human *FGF* genes are scattered throughout the genome, indicating that they were

generated by gene duplications and translocations during evolution [1]. However, several *FGF* genes (despite being from different subfamilies) are clustered. *FGF3*, *FGF4* and *FGF19* are located at 11q13, and are separated by only 40 kb and 10 kb, respectively. *FGF6* and *FGF23* are also located within 55 kb of each other at 12p13. This clustering of these genes supports a model of local gene duplications followed by global gen(om)e duplications.

Genome duplication and the evolution of *Fgf* genes

The evolutionary lineage of multicellular organisms (metazoa), the genome sequences of which are close to completion, is shown in Figure 2. Numerous gene families were expanded in two major phases during evolution (Figure 2). One wave of gene duplication occurred in the metazoan lineage, before the divergence of protostomes and deuterostomes, and another took place in the chordate lineage, during the early emergence of vertebrates [3,10,11]. The *Fgf* gene family also appears to have expanded in two phases. Two or three genes were probably present in primitive metazoans. In the first phase, before chordate evolution, *Fgf* genes were expanded from two or three to six genes by gene duplication (Figure 3a). The most primitive metazoan known to have *Fgf* genes is the nematode, *C. elegans* (Figure 3a). Two *Fgf* genes *egg laying defective 17* (*egl-17*) and *let-756* were identified in *C. elegans* and encode FGF-like proteins of 216 and 425 amino acids, respectively [1]. Three *Fgf* genes, *branchless*, *pyramus* (also known as *fgf8-like2*) and *thisbe* (also known as *fgf8-like1*), have been identified in the fruit fly, *Drosophila melanogaster*. These genes encode large FGF-like proteins of 748–770 amino acids [1,12,13]. Each gene encodes an FGF domain with similarity to the vertebrate FGF core domain. These genes also encode a second region with weaker similarity to the FGF core. It is therefore possible that *Drosophila* FGFs will act as functional FGF dimers (D.M. Ornitz, unpublished).

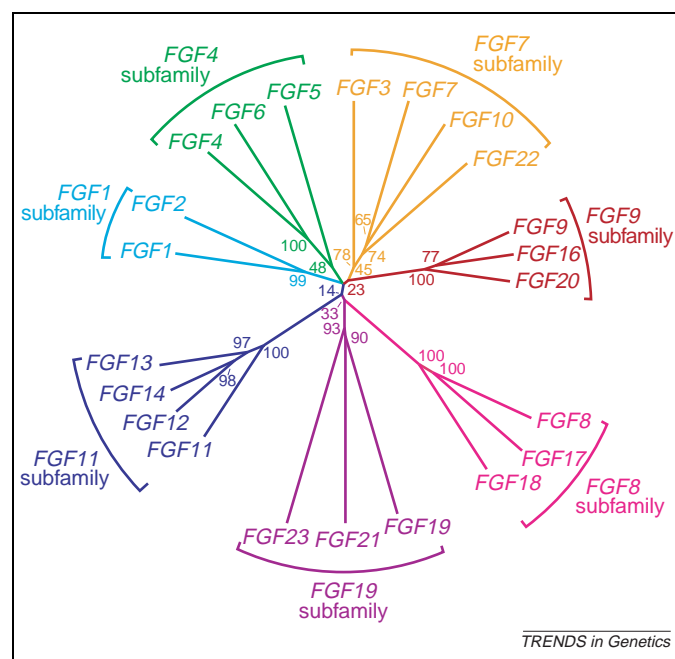


Figure 1. The evolutionary relationships within the human fibroblast growth factor (FGF) gene family. Twenty-two FGF encoding genes have been identified in the human genome. Phylogenetic analyses suggest that these genes can be arranged into seven subfamilies, each containing two to four members. Branch lengths are proportional to the evolutionary distance between each gene. The value at each branch-point indicates the percentage of times that a node was supported in 1000 bootstrap pseudoreplications. Note that *Fgf15* is a mouse gene that is thought to be the ortholog of human *FGF19* [1].

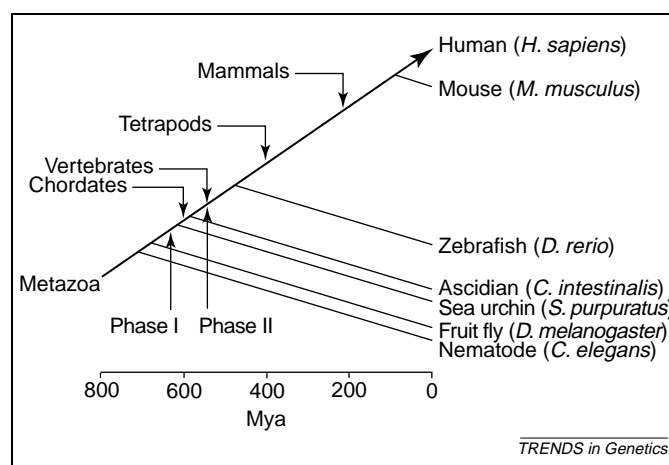


Figure 2. The evolutionary lineage of metazoan organisms. Whole-genome sequences of the nematode, fruit fly, sea urchin, ascidian, zebrafish, mouse and human have been completed or almost completed. Many gene families were expanded in two major phases during evolution (adapted from the lineage by A. Sidow [26]). Phase I occurred in the metazoan lineage, before the divergence of protostomes and deuterostomes. Phase II occurred in the chordate lineage, during the early emergence of vertebrates. In the phase II, two large-scale gene duplications and, possibly, whole-genome duplications, occurred. Abbreviation: Mya, million years ago.

Table 1. Chromosomal localizations of human *FGF* genes and murine *Fgf*^{-/-} phenotypes^a

Gene	Location ^c	<i>Fgf</i> ^{-/-} phenotype (mice)	Description
FGF1 subfamily			
<i>FGF1</i>	5q31.3	Viable	None identified
<i>FGF2</i>	4q27	Viable	Mild cardiovascular, skeletal and neuronal defects
FGF4 subfamily			
<i>FGF4</i>	11q13.3	Lethal, E4–5	Inner cell mass proliferation
<i>FGF5</i>	4q21.21	Viable	Long hair, angora mutation
<i>FGF6</i>	12p13.32	Viable	Subtle, muscle regeneration
FGF7 subfamily			
<i>FGF3</i>	11q13.3	Viable	Mild inner ear, skeletal (tail) and CNS defects
<i>FGF7</i>	15q21.2	Viable	Hair follicle growth, ureteric bud growth
<i>FGF10</i>	5p12	Lethal, P0	Multiple organ development (limb, lung)
<i>FGF22</i>	19p13.3	ND	ND
FGF9 subfamily			
<i>FGF9</i>	13q12.11	Lethal, P0	Lung mesenchyme, heart, gastrointestinal tract, skeleton and testes development
<i>FGF16</i>	Xq21.1 ^d	ND	ND
<i>FGF20</i>	8p22	ND	ND
FGF8 subfamily			
<i>FGF 8</i>	10q24.32	Lethal, E8	Gastrulation defect, CNS and limb development
<i>FGF17</i>	8p21.3	Viable	Cerebellar development
<i>FGF18</i>	5q35.1	Lethal, P0	Skeletal and lung development
FGF11 subfamily			
<i>FGF11</i>	17p13.1	ND	ND
<i>FGF12</i>	3q28	Viable	Neuromuscular phenotype
<i>FGF13</i>	Xq26.3	ND	ND
<i>FGF14</i>	13q33.1	Viable	Neurological phenotype
FGF19 subfamily			
<i>FGF19</i> ^b	11q13.3	Lethal, E12.5	Not clear
<i>FGF21</i>	19q13.33	ND	ND
<i>FGF23</i>	12p13.32	Viable	Growth retardation with abnormal bone phenotype

^aAbbreviations: CNS, central nervous system; ND, not determined.

^bHuman *FGF19* and mouse *Fgf15* are orthologous genes.

^cFor more information, see <http://www.gene.ucl.ac.uk/nomenclature>.

^dHuman *FGF16* was reported to reside on chromosome 8q21 [27]. However, recent analyses suggest that *FGF16* actually lies on the X chromosome (N. Itoh, unpublished). Mouse *Fgf16* is closely associated with *Atrx* (α thalassemia/mental retardation syndrome X-linked homolog) on chromosome X3C. Human *ATRX* is located on chromosome Xq21.1, where there is a large undetermined ~500 kb gap in the genomic sequence. *FGF16* is therefore expected to be located in this region of the X chromosome (N. Itoh, unpublished).

Interestingly, analysis of the ascidian, *Ciona intestinalis*, genome has identified six *Fgf* genes [14]. Five of the six *Fgf* genes, including *Ci-Fgf3/7/10/22*, *Ci-Fgf4/5/6*, *Ci-Fgf8/17/18*, *Ci-Fgf9/16/20* and *Ci-Fgf11/12/13/14*, appear to be derived from genes that are ancestral to the vertebrate *Fgf* subfamilies (Figures 1 and 3a). The sixth member of the *C. intestinalis* *Fgf* family, *Ci-FgfL*, could not be assigned to any vertebrate counterpart.

C. elegans egl-17 has four introns in its coding region. The intron–exon organization of *egl-17* is identical to that of *Ci-Fgf11/12/13/14* and vertebrate *Fgf11–Fgf14* but not to that of other *Ciona* and vertebrate *Fgfs*. Therefore, *egl-17* might be derived from an ancestor of *Ci-Fgf11/12/13/14* and vertebrate *Fgf11–Fgf14* (Figure 3a). By contrast, *C. elegans let-756* has seven introns in its coding region. The intron–exon organization of *let-756* is different from those of vertebrate *Fgf* genes, indicating that a *let-756* ancestral gene might not have evolved into a vertebrate *Fgf* (Figure 3a) and might have been lost during early metazoan evolution. The genes *branchless*, *pyramus* and *thisbe* in *D. melanogaster* and most *Fgf* genes in *C. intestinalis* and vertebrates have three exons encoding their core coding regions, and their intron–exon organization is essentially identical. Of the three *Drosophila* FGF proteins, the core region of

Branchless is most similar to those of most *C. intestinalis* and vertebrate *Fgfs*. These findings indicate that *branchless* but not *pyramus* and *thisbe* might have been derived from an ancestor of chordate *Fgfs* (Figure 3a). Similar to *let-756*, the *pyramus* and *thisbe* ancestral genes might have been lost during early metazoan evolution.

This expanded family of *Fgfs* might be the ancestors of the contemporary vertebrate *Fgf* gene subfamilies. The *Ciona Fgfs*, *Ci-Fgf3/7/10/22*, *Ci-Fgf4/5/6*, *Ci-Fgf8/17/18*, *Ci-Fgf9/16/20* and *Ci-Fgf11/12/13/14* might represent these ancestral subfamilies. During the evolution of early vertebrates, two large-scale gene duplications, possibly whole-genome duplications, occurred generating the full complement of vertebrate genes (Figure 3a). Possible mechanisms of *Fgf* gene evolution are shown in Figure 3a.

Genes closely linked to FGFs on human chromosomes

The human *FGF11* subfamily consists of four members (Figure 1). Examination of the *FGF11–FGF14* gene loci identified close linkage to the claudin genes (*CLDNs7, 16, 2 and 10*) (Figure 4a). This linkage is consistent with two successive whole gen(om)e duplications (Figure 3a). The first and second gen(om)e duplication would produce two and four paralogs, respectively. Phylogenetic analyses

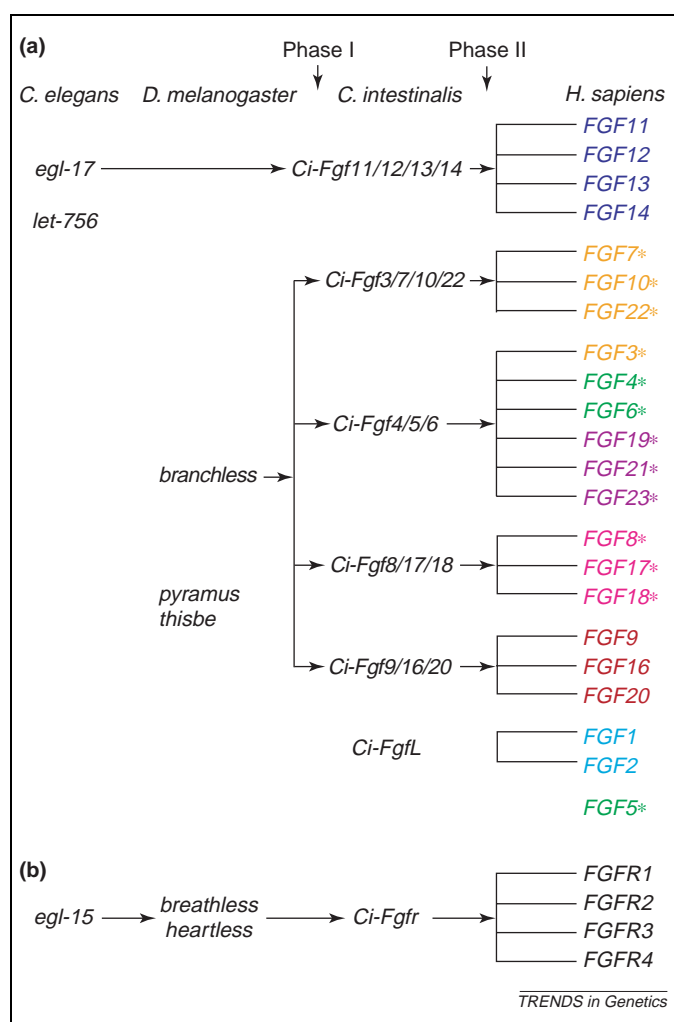


Figure 3. The possible evolutionary relationships between the fibroblast growth factor (FGF) and FGF receptor (FGFR) gene families in different organisms. **(a)** *Caenorhabditis elegans*, *Drosophila melanogaster* and *Ciona intestinalis* have two, three and six *Fgf* genes, respectively. By contrast, *Homo sapiens* has 22 *FGF* genes. The numbers of genes in these organisms is consistent with two phases of gene duplications during metazoan evolution. In phase I, *Fgf* genes were expanded by gene duplication in the metazoan lineage, before the divergence of protostomes and deuterostomes. *C. elegans* *egl-17* might have shared a common ancestral locus with *Ci-Fgf11/12/13/14*. *D. melanogaster* *branchless* might have shared a common ancestral locus with chordate *Fgfs*. Five *Fgf* genes in *C. intestinalis* might have shared a common ancestral locus with vertebrate *Fgf* subfamilies. In the evolution of early vertebrates (phase II), members of the human *FGF* subfamilies were probably generated. The human *FGF11* subfamily with four members might have arisen from a common ancestral gene by two successive gen(om)e duplications. The *FGF7*, *FGF8* and *FGF9* subfamilies with three members might have lost one paralog in some vertebrates after the second vertebrate gen(om)e duplication. *FGFs* 3, 4, 6, 19, 21 and 23 might have arisen from a common ancestral gene by a combination of local gene duplication events, subsequent divergence and then large-scale gen(om)e duplication events. *FGF1* and *FGF2* might have arisen from a common ancestral gene that could not be found in the *Ciona* genome. An ancestral gene of *FGF5* remains to be elucidated. The *FGF* subfamilies are indicated in colored letters according to the subfamilies described in Figure 1. *FGFs* with amino-terminal signal peptides are indicated with asterisks. **(b)** *C. elegans*, *D. melanogaster* and *C. intestinalis* have one or two genes encoding *FGFs*, whereas *H. sapiens* has four genes encoding *FGFRs*. The expansion of the *FGFR* gene family appears to have arisen from only two large-scale gen(om)e duplications during early vertebrate evolution (phase II). The expansion of the *FGFR* gene family was linked to that of *FGF* gene subfamilies in time. The branch lengths are not proportional to sequence divergence.

indicate that the *FGF7* subfamily also consists of four members (Figure 1). Analyses of their gene loci show that *FGFs* 7, 10 and 22, but not *FGF3*, are closely linked to the hyperpolarization-activated, cyclic nucleotide-gated

potassium channel genes (*HCNs4*, 1 and 2) (Figure 4), indicating that *FGFs* 7, 10 and 22 might have arisen from a common ancestral gene by two successive genome duplications (Figure 3a). However, as described in the following section, *FGF3* might have arisen from another ancestral gene. The *FGF8* and *FGF9* subfamilies consist of only three members (Figure 1). *FGFs* 8, 17 and 18 are closely linked to the nucleophosmin genes (*NPMs3*, 2 and 1) (Figure 4a), indicating that these *FGFs* might have arisen from a common ancestral gene (Figure 3a). *FGFs* 9, 16 and 20 are also closely linked to the myotubularin related protein genes (*MTMRs6*, 1 and 7) and zinc finger, DHHC domain containing genes (*ZDHHCs10* and 2) (Figure 4a), indicating that these *FGFs* might also have arisen from a common ancestral gene (Figure 3a). *FGF* subfamilies with three members might have lost one paralog in some vertebrates after the second vertebrate gen(om)e duplication. For example, a new *fgf* gene, *fgf24*, was identified in zebrafish [15]. It appears to be the fourth member of the *fgf8* subfamily. However, an ortholog of *fgf24* has been not identified in the human or mouse gen(om)e and might have been lost during mammalian evolution.

Phylogenetic analyses indicate that *FGF3*, *FGF4* and *FGF6*, and *FGFs19*, 21 and 23 belong to the *FGF7*, *FGF4* and *FGF19* subfamilies, respectively (Figure 1). However, all of these genes are closely linked to the cyclin D genes (*CCNDs1* and 2), potassium voltage-gated channel, shaker-related genes (*KCNAs1*, 6 and 7) and protein tyrosine phosphatase, receptor type, f polypeptide, interacting protein, α genes (*PPFIAs1-3*) (Figure 4a), indicating that these *FGFs* might have arisen from a common ancestral gene by a combination of local gene duplication events, subsequent divergence and then large-scale gen(om)e duplication events (Figure 3a). A potential ancestral gene of the *FGF4* subfamily but not the *FGF19* subfamily was found in the *Ciona* genome (Figure 3a). Therefore, the *FGF19* subfamily might have arisen from an ancestral *FGF4*-related gene by local gene duplication, followed by divergence in amino acid sequence. Phylogenetic analyses and its biochemical function indicate that *FGF3* is a potential member of the *FGF7* subfamily (Figure 1). However, the *FGF3* genomic locus indicates that it might have arisen from *FGF4* by local gene duplication followed by divergent evolution (Figures 3a and 4a). Alternatively, it might have arisen from an ancestor of the *FGF7* subfamily by gene duplication and chance insertion near *FGF4*.

FGF1 and *FGF2* are closely linked to the sprouty genes (*SPRY4* and *SPRY1*) (Figure 4a), indicating that these might have arisen from a common ancestral gene (Figure 3a). However, a potential derivative of an *FGF* ancestral gene could not be found in the *Ciona* genome (Figure 3a) and therefore might have been lost during evolution. Phylogenetic analyses indicate that *FGF5* might have arisen from an ancestral gene of the *FGF4* subfamily (Figure 1). However, the *FGF5* locus shows no apparent linkage to other *FGF* subfamily genes. Therefore, an ancestral *FGF5* gene remains to be elucidated (Figure 3a).

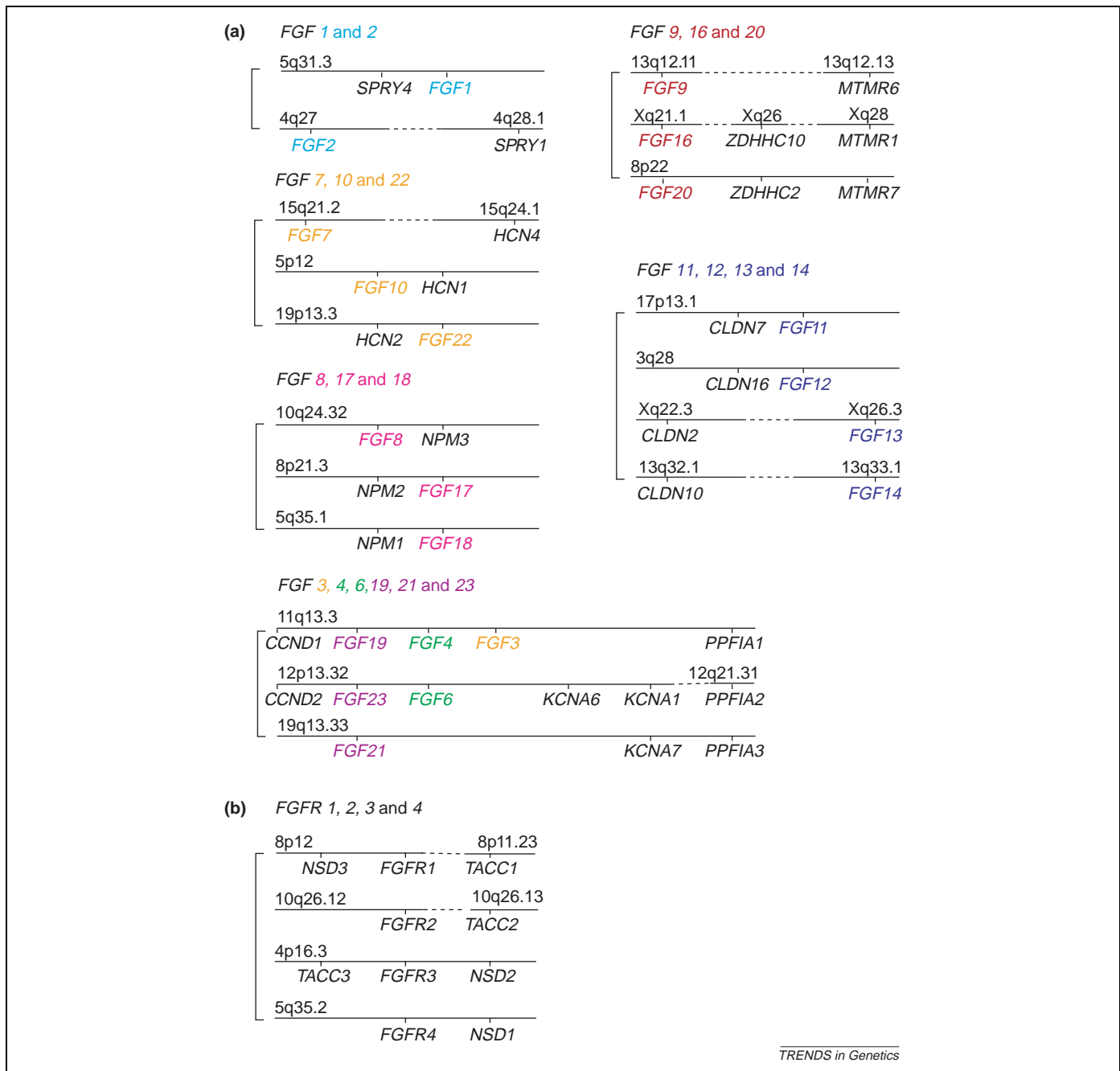


Figure 4. Gene locus maps for fibroblast growth factor (FGF) and FGF receptor (FGFR) genes on human chromosomes. Gene maps were constructed by examining human *FGF* and *FGFR* gene loci using the Ensembl Genome Browser (<http://www.ensembl.org/>) and LocusLink (<http://www.ncbi.nlm.nih.gov/LocusLink/>). **(a)** *FGF* gene loci and closely linked genes are shown. The conservation of gene order supports a model for large-scale gen(om)e duplication events. The *FGF* subfamilies are indicated with colored letters according to Figure 1. **(b)** *FGFR1–FGFR4* gene loci and closely linked genes are shown. The conservation of gene order supports a model for large-scale gen(om)e duplication events. The bar lengths are not proportional to the distances of genes. Abbreviations: *CLDN*, claudin; *HCN*, hyperpolarization-activated, cyclic nucleotide-gated potassium channel; *KCNA*, potassium voltage-gated channel, shaker-related; *MTMR*, myotubularin related protein; *NPM*, nucleophosmin; *NSD*, nuclear receptor-binding SET-domain protein; *PPFIA*, protein tyrosine phosphatase, receptor type, f polypeptide, interacting protein, alpha; *SPRY*, sprouty. *TACC*, transforming, acidic coiled-coil containing protein; *ZDHHC*, zinc finger, DHHC domain containing.

Fgf receptors

Fgfrs are receptor tyrosine kinases (~800 amino acids) that contain an extracellular ligand-binding domain, a transmembrane domain and a split intracellular tyrosine kinase domain. The extracellular region contains two or three immunoglobulin (Ig)-like domains, and a heparin-binding domain [4,16]. Fgfs bind to Fgfrs and induce their dimerization and the phosphorylation of specific cytoplasmic tyrosine residues. The phosphorylation of

Fgfrs triggers the activation of cytoplasmic signal transduction pathways.

The Fgf receptor gene family in vertebrates

Four functional *Fgfr* genes, *Fgfr1–Fgfr4* have been identified in vertebrates including humans, mice and zebrafish. Human *FGFR1–FGFR1* contain 18 coding exons, and have an identical exon–intron organization [16]. These genes are distributed throughout the genome

Table 2. Chromosomal localizations and ligand-specificities of human FGFRs

Gene	Location ^a	Protein ^b	Ligand-specificity ^c
FGFR1	8p12	FGFR1b	FGF1, 2, 3 and 10
		FGFR1c	FGF1, 2, 4, 5 and 6
FGFR2	10q26.12	FGFR2b	FGF1, 3, 7 and 10
		FGFR2c	FGF1, 2, 4, 6 and 9
		FGFR3b	FGF1 and 9
FGFR3	4p16.3	FGFR3c	FGF1, 2, 4, 8 and 9
		FGFR4	FGF1, 2, 4, 6, 8 and 9

^aFor more information see, <http://www.gene.ucl.ac.uk/nomenclature>.

^bAlternative splicing has increased the functional diversity of fibroblast growth factor receptors (FGFRs). Alternatively spliced forms of immunoglobulin (Ig) domains III (the IIIb and IIIc isoforms) of FGFRs are shown.

^cThe ligand-specificities of FGFRs for FGF1–FGF10 are shown [1].

(Table 2), indicating that they were probably also derived from a common ancestral gene by two successive waves of gen(om)e duplication within the chordate lineage. There is no apparent relationship between the chromosomal locations of the genes encoding FGFs and FGFRs.

Evolution of Fgfrs

One *C. elegans* *Fgfr* gene *egg laying defective 15* (*egl-15*) and two *D. melanogaster* *Fgfr* genes (*breathless* and *heartless*) have been identified. *Egl-15* has three Ig domains; *Breathless* and *Heartless* have three and two Ig domains, respectively. These invertebrate *Fgfrs* share only ~30% and ~60% amino acid identity with their vertebrate homologs in the extracellular ligand binding domain and the intracellular kinase domain, respectively. The sea urchin, *Stronglyocentrotus purpuratus* has one *Fgfr* gene (*spFgfr*) that contains three Ig domains [17]. *C. intestinalis* also has only one *Fgfr* gene (*Ci-Fgfr*) that contains two Ig domains [18].

Human *FGFR1–FGFR4* are closely linked to the nuclear receptor-binding SET-domain protein genes (*NSD1–NSD3*) and transforming, acidic coiled-coil containing protein genes (*TACC1–TACC3*) (Figure 4b). These findings indicate that the expansion of the *FGFR* gene family was achieved from a common ancestral gene by only two large-scale gen(om)e duplications during early vertebrate evolution (Figure 3b), and indicates that the expansions of the *FGFR* family and *FGF* subfamilies were linked in time. This co-evolution made increased combinatorial specificities between ligand-receptor pairs possible, and presumably facilitated complex vertebrate evolution.

Alternative splicing has also increased the functional diversity of *Fgfrs*. This process regulates the number (two or three) of Ig domains and the specific sequence of Ig domains III. The ligand-binding properties of *Fgfrs* with two or three Ig domains appear to be similar. By contrast, the ligand-binding specificity and tissue-specific expression properties of alternatively spliced forms of Ig domains III (the IIIb and IIIc isoforms) are different and, consequently, are essential to the functions of the *Fgf* signaling system [4,16]. Phylogenetic analyses of the alternative splicing unit encoding the Ig domain III in *Fgfrs* from *S. purpuratus* to vertebrates show a remarkable conservation of intronic control elements, in both structure and function [17]. These findings indicate that the mechanisms that regulate this alternative splicing

unit have been conserved since the divergence of echinoderms and vertebrates.

Evolution of *Fgf* signaling functions

Major expansions of the *Fgf* family in early vertebrate development enabled their proteins to acquire more important roles in embryonic development [19], tissue repair, metabolism and physiological homeostasis [16]. Many members of the *Fgf* gene family have been disrupted by homologous recombination in mice. Phenotypes range from early embryonic lethality to subtle phenotypes in adult mice (Table 1). Because the *Fgfs* within a subfamily have similar receptor-binding properties and overlapping patterns of expression, functional redundancy is likely to occur. For example, all three members of the *Fgf8* subfamily (*Fgfs* 8, 17 and 18) are expressed in the midbrain–hindbrain junction, and *Fgf* 8 and *Fgf17* cooperate to regulate neuroepithelial proliferation [19]. However, these genes are also expressed in unique locations, and appear to have also evolved independent functions. For example, *Fgf8* is expressed in developing limb ectoderm and is essential for limb bud development. By contrast, *Fgf18* is expressed in developing bone and is essential for skeletal development [20,21]. Redundant functions of *Fgfs* that belong to different subfamilies are also observed. *Fgf2* is apparently the major regulator of the gene encoding glial fibrillary acidic protein (*Gfap*) expression because in mice lacking *Fgf2*, GFAP is distinctly reduced in the cortex and striatum, whereas in *Fgf5* null animals only a reduction in the midbrain tegmentum can be observed. However, in *Fgf2/Fgf5* double null-mutant animals, GFAP-immunoreactivity is reduced in all three-brain regions [22]. In zebrafish, *fgf3* and *fgf8* also encode redundant genes that are required for otic placode induction and forebrain development [23,24].

Concluding remarks

The *Fgf* signaling system has been identified in multicellular organisms ranging from *C. elegans* to vertebrates but not in unicellular organisms. The vertebrate *Fgf* and *Fgfr* gene families consist of 22 and four members, respectively. The vertebrate *Fgf* gene family appears to have expanded in two phases. In the first phase, before the divergence of the chordate lineage, five vertebrate *Fgf* gene subfamilies were generated from ancestral *Fgf* genes by gene duplication. Three or four members of the vertebrate subfamilies were generated by two large-scale gen(om)e duplications during the evolution of early vertebrates. Four members of the vertebrate *Fgfr* gene family were also generated during this phase. This co-evolution has permitted the evolution of increased ligand-receptor specificity, enabling the formation of preferred ligand-receptor interactions.

Alternative splicing has increased the functional diversity of the *Fgf* signaling system by regulating ligand-binding specificity through tissue-specific alternative splicing of Ig domain III of *Fgfr1–Fgfr3*. The mechanisms that regulate this alternative splicing unit have been conserved since the divergences of echinoderms and vertebrates, before the expansion of the *Fgfr* gene family. Ligand-receptor specificity has also been refined by

the evolution of glycosaminoglycan cofactors that modulate ligand availability and receptor binding [4,25]. Evolution of primitive metazoan organisms to vertebrates required an increase in the complexity of developmental signaling systems. A significant part of this capacity for increased signaling complexity appears to have resulted from the expansion of the *Fgf* and *Fgfr* gene families.

Acknowledgements

We thank Morichika Konishi and Diane Redmond for assistance with graphics and Jeremy Gibson-Brown for critically reading this manuscript. This work was supported by the Grant-in Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology, Japan and by a grant from the Takeda Science Foundation, Japan to N.I. and grants from the National Institutes of Health and March of Dimes to D.M.O.

References

- Ornitz, D.M. and Itoh, N. (2001) Fibroblast growth factors. *Genome Biol.* 2, 3005.1–3005.12 (<http://genomebiology.com/2001/2/3/reviews/3005>)
- Ensoli, B. *et al.* (2003) The fibroblast growth factors. In *The Cytokine Handbook* (Thomson, A.W. and Lotze, M.T. eds), pp. 748–781, Academic Press
- Thornton, J.W. and DeSalle, R. (2000) Gene family evolution and homology: genomics meets phylogenetics. *Annu. Rev. Genomics Hum. Genet.* 1, 41–73
- Ornitz, D.M. (2000) FGFs, heparan sulfate and FGFRs: complex interactions essential for development. *BioEssays* 22, 108–112
- Tassi, E. *et al.* (2001) Enhancement of fibroblast growth factor (FGF) activity by an FGF-binding protein. *J. Biol. Chem.* 276, 40247–40253
- Schoorlemmer, J. and Goldfarb, M. (2001) Fibroblast growth factor homologous factors are intracellular signaling proteins. *Curr. Biol.* 11, 793–797
- Rat Genome Sequencing Project Consortium. (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428, 493–521
- Aparicio, S. *et al.* (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* 297, 1301–1310
- Nusslein-Volhard, C. *et al.* (2002) Zebrafish as a system to study development and organogenesis. In *Zebrafish* (Nusslein-Volhard, C. and Dahm, R. eds), pp. 1–6, Oxford University Press
- Sidow, A. (1996) Gen(om)e duplications in the evolution of early vertebrates. *Curr. Opin. Genet. Dev.* 6, 715–722
- Horton, A.C. *et al.* (2003) Phylogenetic analyses alone are insufficient to determine whether genome duplication(s) occurred during early vertebrate evolution. *J. Exp. Zool. Part B. Mol. Dev. Evol.* 229, 41–53
- Stathopoulos, A. *et al.* (2004) pyramus and thisbe: FGF genes that pattern the mesoderm of *Drosophila* embryos. *Genes Dev.* 18, 687–699
- Gryzik, T. and Muller, H.-A.J. (2004) *FGF8-like1* and *FGF8-like2* encode putative ligands of the FGF receptor Htl and are required for mesoderm migration in the *Drosophila* gastrula. *Curr. Biol.* 14, 659–667
- Satou, Y. *et al.* (2002) Fgf genes in the basal chordate *Ciona intestinalis*. *Dev. Genes Evol.* 212, 432–438
- Draper, B.W. *et al.* (2003) Zebrafish *fgf24* functions with *fgf8* to promote posterior mesodermal development. *Development* 130, 4639–4654
- Powers, C.J. *et al.* (2000) Fibroblast growth factors, their receptors and signaling. *Endocr. Relat. Cancer* 7, 165–197
- Mistry, N. *et al.* (2003) Of urchin and men: Evolution of an alternative splicing unit in fibroblast growth factor receptor genes. *RNA* 9, 209–217
- Satou, Y. *et al.* (2003) A genomewide survey of developmentally relevant genes in *Ciona intestinalis*. V. Genes for receptor tyrosine kinase pathway and Notch signaling pathway. *Dev. Genes Evol.* 213, 254–263
- Xu, J.S. *et al.* (2000) Temporal and spatial gradients of Fgf8 and Fgf17 regulate proliferation and differentiation of midline cerebellar structures. *Development* 127, 1833–1843
- Liu, Z. *et al.* (2002) Coordination of chondrogenesis and osteogenesis by fibroblast growth factor 18. *Genes Dev.* 16, 859–869
- Ohbayashi, N. *et al.* (2002) FGF18 is required for normal cell proliferation and differentiation during osteogenesis and chondrogenesis. *Genes Dev.* 16, 870–879
- Reuss, B. *et al.* (2003) Functions of fibroblast growth factor (FGF)-2 and FGF-5 in astroglial differentiation and blood-brain barrier permeability: evidence from mouse mutants. *J. Neurosci.* 23, 6404–6412
- Liu, D. *et al.* (2003) Fgf3 and Fgf8 dependent and independent transcription factors are required for otic placode specification. *Development* 130, 2213–2224
- Walshe, J. and Mason, I. (2003) Unique and combinatorial functions of Fgf3 and Fgf8 during zebrafish forebrain development. *Development* 130, 4337–4349
- Allen, B.L. and Rapraeger, A.C. (2003) Spatial and temporal expression of heparan sulfate in mouse development regulates FGF and FGF receptor assembly. *J. Cell Biol.* 163, 637–648
- Sidow, A. (2002) Sequence first. Ask questions later. *Cell* 111, 13–16
- Kim, H.-S. (2001) The human FGF gene family: chromosome location and phylogenetic analysis. *Cytogenet. Cell Genet.* 93, 131–132