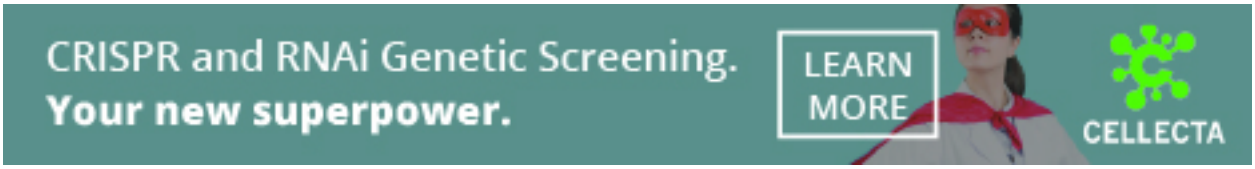
G protein alpha subunit multigene family in the Japanese puffer fish Fugu rubripes: PCR from a compact vertebrate genome.

M M Sarwal, J M Sontag, L Hoang, et al.

Genome Res. 1996 6: 1207-1215
Access the most recent version at doi:10.1101/gr.6.12.1207



LETTER

G Protein Alpha Subunit Multigene Family in the Japanese Puffer Fish *Fugu rubripes*: PCR from a Compact Vertebrate Genome

Minnie M. Sarwal,[1] Jean-Marie Sontag,[2] Lieu Hoang,[2] Sydney Brenner,[1] and Thomas M. Wilkie[2,3]

[1]Molecular Genetics Unit, Department of Medicine, Addenbrooke's Hospital, Cambridge, UK;
[2]Pharmacology Department, Southwestern Medical Center, University of Texas, Dallas, Texas 75235-9041

We compare the complexity and organization of the G protein α subunit multigene family in the vertebrate genomes of mammals and the Japanese puffer fish *Fugu rubripes*. Fourteen *Fugu* Gα genes were identified of the 16 genes characterized previously in mammals, including *Fugu* genes from the four classes of alpha subunits Gs, Gi, Gq, and G12. *Fugu* and mammalian Gα coding sequences are highly homologous, and the intron/exon structure of the fish and mammalian orthologs is identical throughout the coding regions. A novel Gα gene, Gαpl, was also identified in *Fugu rubripes* and two other species of puffer fish. The complete sequence of *Gnaz* and the tandemly duplicated genes *Gnai2* and *Gnat1* were obtained from a *Fugu* genomic cosmid library. Introns in the puffer fish Gα genes lacked repeat DNA sequences, other than simple sequence length repeats, and most introns were significantly shorter in *Fugu* than in mammalian orthologs. The compact genome of puffer fish provides a unique vertebrate model for characterizing multigene families and identifying novel genes directly from genomic DNA by PCR amplification with degenerate primers. The fact that *Fugu* encodes most, if not all, of the G protein alpha subunits identified in mammals strongly supports *Fugu* as a model organism for vertebrate genome research.

The sequence data described in this paper have been submitted to the GenBank data library under accession nos. *Gnai1* (L78800, L79903), *Gnai2* (L79898), *Gnai3* (L79894), *Gnai2-like* (L79889, L79890), *Gnao* (L79891, 8L79892), *Gnat1* (L79908), *Gnat2* (L79897, L79902), *Gnaz* (L79900, L79901), *Gnaq* (L79896, L79904, L79905), *Gna11* (L79906), *Gna14* (L78803), *Gnapl* (L79895), *Gna12* (L79907), *Gna13* (L78931), *Gnas* (L78839, L79893), *Gnas-like* (L78929, L78930), and *Gnal1* (L78919).

Heterotrimeric G proteins, composed of α, β, and γ subunits, are present in all higher eukaryotes, including fungi, plants, and metazoans (Simon et al. 1991), and coordinate various metabolic, humoral, neural, and developmental functions (Gilman 1987; Birnbaumer 1990). Metazoan organisms uniquely express four classes of α subunits, Gs, Gi, Gq, and G12, based on sequence conservation and functional similarities (Wilkie et al. 1992). Members of the Gq class activate PLCβ isoforms; Gs members activate adenylyl cyclases; and Gi members have several effector targets, but almost all share the common feature that they are sensitive to ADP ribosylation by pertussis toxin, and G12 class proteins influence Rho-dependent formation of stress fibers (Buhl et al. 1995). The repertoire of Gα subunits has been characterized extensively in mouse, human, and two invertebrates, *Drosophila melanogaster* and *Caenorhabditis elegans*. Within the four classes of Gα subunits, mammals express several related genes with similar receptor and effector specificities. In total, mammals have at least 16 functional Gα genes, several of which are spliced alternatively, that encode 20 distinct protein products (Wilkie and Yokoyama 1994). The numerous closely related mammalian genes, such as the tandemly duplicated genes *Gnai2/Gnat1* and *Gnai3/Gnat2*, apparently arose through multiple independent gene duplication events subsequent to vertebrate and invertebrate divergence (Wilkie et al. 1992). In contrast, flies and worms apparently encode only one obvious ortholog of the

[3]Corresponding author.
E-MAIL wilkie@utsw.swmed.edu; FAX 214-648-2994.

closely related mammalian Gα genes, although some of the invertebrate genes, such as *Drosophila dgq*, are also spliced alternatively (Talluri et al. 1995). Invertebrates also express several distantly related Gα genes that have not been identified in mammals, despite concerted attempts to clone them, such as *Drosophila dgf* (Quan et al. 1989), snail Gαa (Knol et al. 1995), and nematode *gpa1*, *gpa2*, and *gpa3* (Lochrie et al. 1991). The question remains whether these divergent genes have evolved specifically in these invertebrates or represent conserved Gα subunits that might couple distinct signaling pathways in both invertebrates and vertebrates.

Teleosts, which include the puffer fish *Fugu rubripes*, are the most extant vertebrates precursors of mammals, with a separation time of 430 million years (Powers 1991). The puffer fish *Fugu rubripes* (*Fugu*) is a useful model for comparative study of vertebrate genomes because of its relatively compact genome, ~7.5 times smaller than mammalian genomes (Brenner et al. 1993). To date, the coding sequences of most *Fugu* genes that have been analyzed are highly homologous to their vertebrate orthologs (Baxendale et al. 1995; Elgar et al. 1995; Macrae and Brenner 1995; Mason et al. 1995), but *Fugu* genes usually have much smaller introns and an astounding paucity of repeat sequences, resulting in a high density of genes per unit length on the chromosome. These physical features facilitate genomic mapping and gene sequencing, and also make *Fugu* genomic DNA an excellent template for PCR amplification of multigene families with degenerate oligonucleotide primers (Macrae and Brenner 1995).

To extend our analysis of comparative gene organization in *Fugu*, mouse, and human, and to explore complexity of the G protein multigene family, we PCR amplified *Fugu* genomic DNA with degenerate oligonucleotide primers that were biased toward the G protein alpha subunits. In this report, we show that the compact *Fugu* genome is an excellent unit copy template for PCR amplification of the Gα multigene family with degenerate primers. We have characterized the entire genomic organization of three Gi class genes, *Gnai2*, *Gnat1*, and *Gnaz*, and find that they are highly homologous to their human orthologs. The order of genes on the chromosome is also conserved between *Fugu* and humans in the immediate vicinity of the Gα genes that we have characterized. Conservation of gene clusters in *Fugu* with their human counterparts may greatly assist the analysis of gene function and the identification of disease loci.

RESULTS

Heterotrimeric G protein alpha subunit genes contain five highly conserved amino acid motifs, referred to as G boxes (Fig. 1), that provide excellent targets for PCR amplification of cDNA with degenerate primers (Wilkie et al. 1994). However, mammalian genomic DNA is not a suitable template for PCR amplification with these degenerate primers (data not shown). Therefore, to explore the diversity of the Gα multigene family on a unit copy template, we used degenerate primers to PCR amplifiy from the condensed genome of the puffer fish *Fugu rubripes*. To limit the length of the expected PCR products from *Fugu* genomic DNA, we used primer pairs that were complementary to adjacent G boxes known to be separated by a single intron in other vertebrate orthologous genes (Fig. 1). Primer pairs were chosen that were biased toward a certain class of Gα subunits. PCR amplification of *Fugu* genomic DNA with the primers iMP19 and TW8 that target the G3 and G4 boxes, respectively, reproducibly generated four bands that included five different Gα genes from the Gi and Gq classes: *Gnai1*, *Gnao*, *Gnaq*, *Gna14*, and a novel sequence called *Gnap1* (Fig. 2). Primer pairs that were biased toward the coding sequence of Gs class genes within the G3 and G4 boxes predominantly amplified *Gnas*, a *Gnas-like* gene, and *Gna1* (encoding

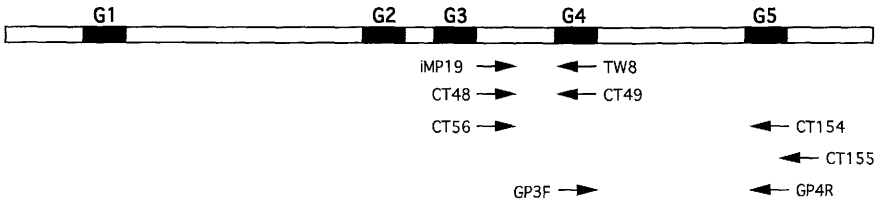


Figure 1 Name and location of degenerate PCR primers to amplify G protein alpha subunit genes. Approximate location is shown of the most highly conserved amino acid motifs in the G alpha concensus sequence, labeled G1 through G5, and the degenerate oligonucleotide pairs that were used in the PCR are shown. Sense and antisense primers are indicated by left and right facing arrows, respectively.

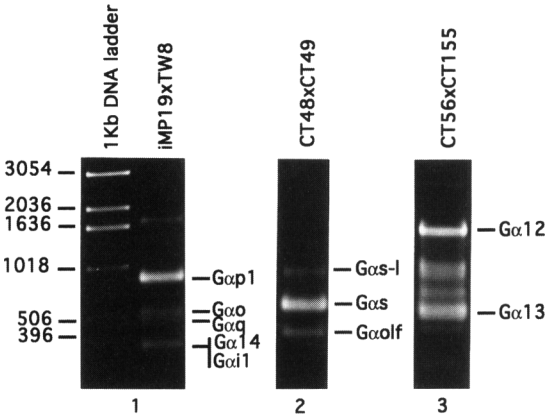


Figure 2 PCR products amplified from *Fugu* genomic DNA with degenerate primer pairs were used in separate reactions: iMP19 and TW8 (panel *1*); CT48 and CT49 (panel *2*); CT56 and CT155 (panel *3*). The G protein alpha subunits that are the predominant PCR products are indicated to the right of each panel, unlabeled PCR fragments did not contain Gα subunit sequences. Markers (1 kb ladder) are shown in panel *1*.

Gαolf); oligonucleotides biased toward the G12 class amplified *Gna12* and *Gna13* (Fig. 2). In separate reactions, we also amplified *Fugu* genomic DNA with degenerate primers that complemented sequence within the G4 and G5 boxes (GP3F and GP4R; Fig. 1). Translation of the *Fugu* gene-coding sequences (Fig. 3) and comparison to mammalian Gα genes showed remarkable conservation across the PCR amplified region (Table 1). The intron/exon boundaries were also conserved in the *Fugu* and mammalian orthologs (Fig. 4). Importantly, Gα subunit genes from each of the four classes were identified by PCR amplification from *Fugu* genomic DNA.

The repertoire of *Fugu* Gα genes was also surveyed by low-stringency hybridization of a gridded *Fugu* genomic cosmid library with Gα probes obtained from PCR amplification of *Fugu* genomic DNA or mammalian cDNAs. The G protein α subunits contained on these cosmids were identified as *Gnai3* and *Gnat2* by DNA sequence of cloned fragments obtained either from random *Sau3A* digestion or PCR amplification with degenerate primers (Figs. 3 and 4). The *Fugu* genes identified from cosmid clones include *Gnas*, *Gna11*, *Gnaq*, *Gnaz*, *Gnat1*, *Gnat2*, *Gnai1*, *Gnai2*, *Gnai3*, and another gene closely related to *Gnai2* termed *Gnai2-like* (Fig. 4). Using the combined approaches of library screens and direct PCR amplification of *Fugu* genomic DNA, we have identified three new Gα genes, *Gnai2-like*, *Gnas-like*, and *Gnap1* that are probably members of the Gi, Gs, and Gq classes, respectively, and 14 of the 16 Gα genes found previously in mammals (Fig. 4).

To compare the structure of *Fugu* and mammalian Gα genes, we cloned and characterized the complete coding sequence of three *Fugu* orthologs of *Gnai2*, *Gnat1*, and *Gnaz*, respectively (Fig. 4). These genes were characterized because the genomic structure of the mammalian genes from several species is known, and *Gnaz*, although also a member of the Gi class, has a very different intron organization than either *Gnai2* or *Gnat1* (Kaziro et al. 1991). Furthermore, *Gnai2* and *Gnat1* cosegregate in mouse and have been shown to be duplicated tandemly in humans, whereas *Gnaz* mapped as a single gene well separated on human chromosome 22q11 from other Gα genes (Wilkie et al. 1992). The amino acid sequence of *Fugu Gnai2*, *Gnat1*, and *Gnaz* are 89%, 95%, and 95% identical to their respective mammalian orthologs, and exhibit all of the most important sequence features that have been characterized in mammalian Gi class α subunits (Simon et al. 1991). As in humans, *Fugu Gnaz* is separate from other Gα genes. In contrast, *Fugu*

Class	Gene	cDNA	
Gq	*Gnaq*	α_q	ENV**T**SIM**F**LVAL**S**E**Y**DQVLV**E**SDNEN**R**MEE**S**KALF**R**TIITYPWFQNS**S**VIL
	Gna14	α_{14}	ENV**T**SII**F**LVAL**S**E**Y**DQVLS**E**CDNEN**R**MEE**S**KALF**K**TIITYPWFQR**S**VIL
	Gnap1	α_{p1}	ENV**T**SLI**F**LASL**S**E**Y**DQVLE**E**RETIN**R**MHE**S**LALF**Y**TTIHSPWFQNT**S**IIL
Gi	*Gnai1*	α_{i1}	EGV**T**AII**F**CVAL**S**D**Y**DLVLA**E**DEEVN**R**MHE**S**MKLF**D**SICNNKWFTDT**S**IIL
	Gnai2	α_{i2}	EGV**T**GII**F**CVGM**S**A**Y**DLVLA**E**DEEMN**R**MHE**S**MKLF**D**SICNNKWFTET**S**IIL
	Gnao	α_o	EDV**T**AII**F**CVAL**S**G**Y**DQVLH**E**DETTN**R**MHE**S**LMLF**D**SICNNKFFIDT**S**IIL
	Gnaz	α_z	EGV**T**AII**F**CVEL**S**G**Y**DLKLY**E**DNQTS**R**MAE**S**LRLF**D**SICNNNWFINT**S**LIL
	Gnat1	α_{t1}	EGV**T**CIE**F**IAAL**S**A**Y**DMVLV**E**DDEVN**R**MHE**S**LHLF**N**SICNHRYFAAT**S**IVL
G12	*Gna12*	α_{12}	DGI**T**SIL**F**MVSS**S**E**Y**DQVLM**E**DRRTN**R**LVE**S**MNIF**E**TIVNNKLFLNV**S**IIL
	Gna13	α_{13}	DSV**T**SIL**F**LVSS**S**E**Y**DQVLM**E**DRQTN**R**LRE**S**VDIF**E**TIVNNRVFGNV**S**IIL
Gs	*Gnas*	α_s	NDV**T**AII**F**VVAS**S**S**Y**NMVIR**E**DNQTN**R**LQE**A**LNLF**K**NIWNNRWLRTI**S**VIL
	Gnal	α_{olf}	NDV**T**AII**F**VAAS**S**S**Y**NMVIR**E**DNSTN**R**LRE**S**LDLF**R**SIWTNRFLKTI**S**VIL
	Gnas-l	α_{s-l}	NDV**T**AII**F**AVAS**S**S**Y**NMVIR**E**DNQTN**R**LQE**A**LNLF**K**NIWNNRWLRTI**S**DIL
	Concensus		**T F S Y E R E F S L**

Figure 3 Predicted amino acid sequence from *Fugu* G protein alpha subunits. The amino acid sequences were compiled from the sequence of all *Fugu* clones between the iMP19 and TW8 primers. Sequences are grouped according to Gα subunit class. Most sequences could be clearly identified, with the exception of Gαq or Gα11 (αq/11) and Gαi1 or Gαi3 (αi1/3). The consensus amino acids, defined by amino acid identity in every sequence, are shown below. Boldface type in the alpha subunit sequences indicates agreement with the consensus.

Table 1. Amino Acid Identifiy of *Fugu* and Mammalian Gα Genes between the Degenerate PCR primers iMP19 and TW8

	Fr Gnai1[a]	Fr Gnai2	Fr Gnao	Fr Gnaz	Fr Gnat1	Fr Gnaq	Fr Gna14
Hs Gnai1[b]	**98.0**	90.2	80.4	70.6	64.7	52.9	54.9
Hs Gnai2	96.1	**92.2**	80.4	70.6	66.7	52.9	54.9
Hs Gnai3	96.1	**92.2**	78.4	70.6	64.7	52.9	54.9
Mm GnaoA[c]	80.4	72.5	**100.0**	72.5	62.7	52.9	54.9
Mm GnaoB	86.3	78.4	94.1	72.5	62.7	54.9	56.9
Hs Gnaz	70.6	66.7	72.5	**100.0**	54.9	49.0	49.0
Hs Gnat1	66.7	62.7	62.7	54.9	**94.1**	49.0	49.0
Mm Gnat2	72.5	68.6	70.6	58.8	92.2	51.0	51.0
Rn Gnag[d]	72.5	68.6	68.6	58.8	92.2	51.0	51.0
Mm Gnaq	52.9	49.0	52.9	49.0	51.0	**100.0**	**90.2**
Mm Ga11	52.9	49.0	52.9	49.0	51.0	**100.0**	**90.2**
Mm Gna14	56.9	52.9	54.9	51.0	49.0	90.2	**90.2**
Mm Gna15	47.1	43.1	49.0	49.0	45.1	62.7	64.7
Hs Gna16	47.1	43.1	49.0	51.0	43.1	58.8	60.8
Mm Gna12	52.9	52.9	52.9	47.1	43.1	49.0	47.1
Mm Gna13	47.1	47.1	52.9	49.0	45.1	51.0	49.0
Hs Gnas	51.0	47.1	54.9	52.9	45.1	41.2	45.1
Hs Gna1	51.0	47.1	54.9	49.0	49.0	39.2	41.2

	Fr Gnap1	Fr Gna12	Fr Gna13	Fr Gras	Fr Gnaolf	Fr Gnas-1
Hs Gnai1[b]	54.9	52.9	49.0	49.0	47.1	49.0
Hs Gnai2	54.9	52.9	49.0	49.0	47.1	49.0
Hs Gnai3	54.9	52.9	49.0	49.0	47.1	49.0
Mm GnaoA[c]	58.8	52.9	52.9	52.9	54.9	52.9
Mm GnaoB	60.8	52.9	52.9	54.9	52.9	54.9
Hs Gnaz	49.0	47.1	49.0	51.0	47.1	51.0
Hs Gnat1	51.0	41.2	43.1	45.1	49.0	45.1
Mm Gnat2	52.9	45.1	43.1	43.1	49.0	43.1
Rn Gnag[d]	52.9	45.1	43.1	45.1	49.0	45.1
Mm Gnaq	**66.7**	49.0	52.9	41.2	43.1	39.2
Mm Ga11	**66.7**	49.0	52.9	41.2	43.1	39.2
Mm Gna14	64.7	51.0	54.9	43.1	45.1	41.2
Mm Gna15	**66.7**	37.3	43.1	37.3	39.2	35.3
Hs Gna16	**66.7**	37.3	43.1	39.2	41.2	37.3
Mm Gna12	47.1	**98.0**	80.4	45.1	41.2	45.1
Mm Gna13	51.0	80.4	**90.2**	51.0	45.1	51.0
Hs Gnas	35.3	45.1	49.0	**98.0**	82.4	**94.1**
Hs Gna1	39.2	43.1	51.0	80.4	**84.3**	78.4

Gene symbols, see Wilkie et al. (1991).
[a]Fr: *Fugu rubripes.*
[b]Hs: *Homo sapiens.*
[c]Mm: *Mus musculus.*
[d]Rn: *Ratus norvegicus.* Gnag encodes gustducin (McLaughlin et al. 1992).

Gnai2 and *Gnat1* are duplicated tandemly (cosmid 27E8; M.M. Sarwal and S. Brenner, unpubl.). *Fugu Gnai2* is the upstream gene in this pair, whereas human *Gnat1* is in the upstream position of a head-to-tail array with *Gnai2* (Sekido et al. 1996). Interestingly, *Fugu Gnai3* and *Gnat1* are also duplicated tandemly but in a tail-to-tail orientation, indicating considerable gene rearrangement since the emergence of the three *Gnai/Gnat* genes in vertebrates.

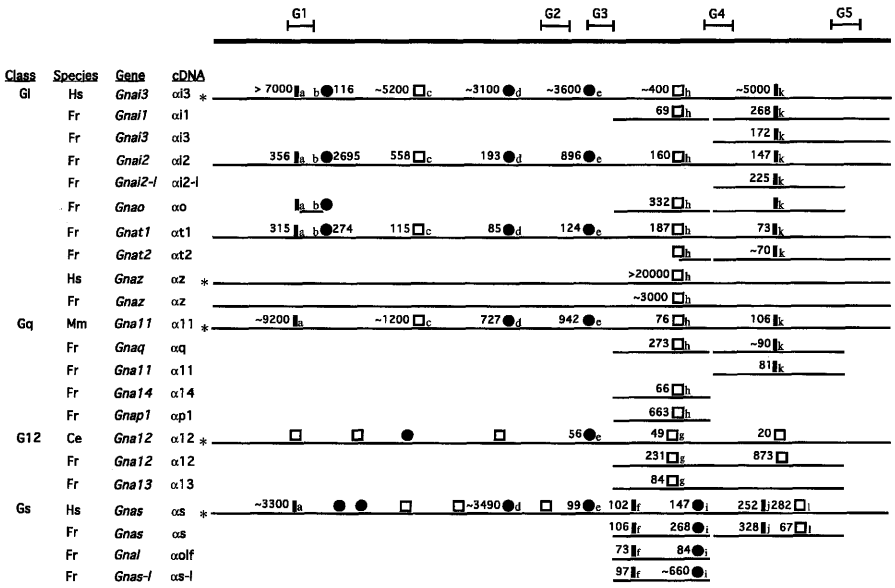


Figure 4 *Fugu* Gna gene structure. The consensus sequence of representative genes from each Gα class was obtained from amino acid alignment (Wilkie and Yokoyama 1994). The consensus sequence is depicted by a thick solid line showing the relative position of amino acids within the GTP-binding domains G1 through G5. Each gene is identified by class, species, gene, and name of cDNA. Thin solid lines depict individual genes, the extent of the line indicates the length and location of sequence obtained for each *Fugu* gene relative to the consensus sequence. The reference genes for each class are indicated (*) to the left of the line; *Gnai3* (Kaziro et al. 1991), *Gnaz* (Matsuoka et al. 1988), *Gna11* (Davignon et al. 1996), *Gna12* (accession no. U39855), and *Gnas* (Kozasa et al. 1988). Intron positions are indicated above each gene. Introns that are positioned after the first, second, or third nucleotide within a codon are identified by a thick vertical line, solid circle, or open square, respectively. The letters a through l identify those introns that occupy exactly the same position in the alignment in more than one gene. Introns of unique position within the alignment are not lettered. The number of nucleotides in *Fugu* and mammalian introns are indicated for comparison. CA repeat sequences were identified in *Fugu Gna12*, intron 7, and *Gnas-like*, intron 9. Accession nos. for the *Fugu* Gα gene sequences are *Gnai1* (L78800, L79903), *Gnai2* (L79898), *Gnai3* (L79894), *Gnai2-like* (L79889, L79890), *Gnao* (L79891, 8L79892), *Gnat1* (L79908), *Gnat2* (L79897, L79902), *Gnaz* (L79900, L79901), *Gnaq* (L79896, L79904, L79905), *Gna11* (L79906), *Gna14* (L78803), *Gnap1* (L79895), *Gna12* (L79907), *Gna13* (L78931), *Gnas* (L78839, L79893), *Gnas-like* (L78929, L78930), and *Gna1* (L78919).

Intron position of *Fugu Gnai2*, *Gnat1*, and *Gnaz* was determined from the genomic DNA sequence across the entire coding region of each gene. The position of all seven introns in *Gnai2* and *Gnat1* is identical in *Fugu* and mammalian orthologs. In contrast to the structure of *Gnai2* and *Gnat1*, mammalian and *Fugu Gnaz* genes have a single intron at a conserved position in the coding sequence (Fig. 4). All introns that we have identified in *Fugu* Gα genes, including portions of nine other genes from the Gi, Gq, and Gs classes, are conserved in number and placement with their mammalian orthologs (Fig. 4). Introns of the *Fugu Gna12* and *Gna13* genes were compared with a *C. elegans* gene (mammalian G12 class genes have not been characterized); the position of one of these introns is unique to the G12 class (intron g; Fig. 4) and placed identically in all three genes, whereas the downstream intron was not present in *Fugu Gna13*, and is in slightly different locations in the *Fugu* and *C. elegans Gna12* genes (Fig. 4) and the *Drosophila* G12 class gene *concertina* (Wilkie and Yokoyama 1994; M. Wayne, pers. comm.). In addition, several mammalian Gα genes are spliced alternatively, including *Gnas* (Kozasa et al. 1988). *Fugu Gnas* encodes two exons that may be spliced alternatively (accession nos. in Fig. 4 legend), analogous to the Gsα-1 and Gsα-3 spliced forms of human *Gnas* (Bray et al. 1986). We have also identified CA repeats, which generate simple sequence length polymorphism (SSLP) in humans and in *Fugu* genes *Gna12* and *Gnas-like* (Fig. 4 legend). The major difference between *Fugu* and mammalian Gα genes is that most introns were significantly shorter in *Fugu* than in their mammalian orthologs (Fig. 4) and the *Fugu* introns did not contain highly repetitive DNA.

DISCUSSION

Metazoan organisms express four classes of heterotrimeric G protein alpha subunits, known as Gs, Gi, Gq, and G12, that are not found among the repertoire of Gα genes in other eukaryotes such as plants, fungi, or *Dictyostelium*. Within metazoa, the amino acid sequence and gene

structure of at least one of the Gα genes within each class is highly conserved, but the number of paralogs within each class and their rate of divergence differs between mammals and invertebrates (Wilkie and Yokoyama 1994; Davignon et al. 1996). The functional significance of these differences is obscured by the evolutionary distance that separates vertebrates and invertebrates. Indeed, it is difficult to discern whether *Drosophila dgf*, snail Gαa, and the *C. elegans* genes *gpa-1*, *gpa-2*, and *gpa-3* have vertebrate orthologs, and attempts to identify these genes in mammals have failed. Therefore, we sought to explore diversity of the Gα multigene family in a model organism more closely related to mammals. Teleosts, which include the puffer fish *Fugu rubripes*, are the most extant vertebrate precursors of mammals, with a separation time of 430 million years (Powers 1991). The genome of *Fugu* is ~7.5 times smaller than human (Brenner et al. 1993). The *Fugu* genome probably accommodates the same complement of genes as humans, but is more compact, with fewer repeat sequences, smaller introns, and shorter intragenic distances. The minimal complexity of the *Fugu* genome allows the rapid isolation of *Fugu* genes by PCR (Macrae and Brenner 1995) and their complete analysis at the genomic level.

To explore further the complexity of the G protein alpha subunit multigene family in vertebrates and to test the utility of *Fugu* as a model vertebrate genome, we used the combined approach of low-stringency screens of genomic cosmid libraries and PCR amplification of *Fugu* genomic DNA with degenerate oligonucleotide primers (Figs. 1 and 2) to clone and characterize members of the G alpha multigene family in *Fugu* (Fig. 3). Using this combined approach, we identified 14 of 16 known mammalian homologs of the Gα multigene family in *Fugu rubripes*. The coding sequences of *Fugu* and mammalian Gα orthologs are highly conserved (Table 1) and their intron/exon boundaries are identical (Fig. 4). Remarkable conservation of coding sequence and gene structure has also been seen between other *Fugu* and mammalian genes (Baxendale et al. 1995; Elgar et al. 1995; Macrae and Brenner 1995; Mason et al. 1995), a property that facilitates the rapid identification of *Fugu* and mammalian homologs.

Comparison of the coding sequence of the Gα genes in *Fugu* with other published vertebrate and invertebrate sequences demonstrates that *Fugu* expresses at least two genes each from the Gi, Gs, Gq, and G12 class (Table 1). We characterized the complete coding sequence and gene structure of three *Fugu* genes in the Gi class, *Gnai2*, *Gnat1*, and *Gnaz*. The amino acid sequences of the *Fugu* genes are very similar to their human and mouse homologs, including conservation of the cysteine four residues from the carboxyl terminus that is ADP ribosylated by pertussis toxin, the consensus sequence for amino myroistoylation (Mumby et al. 1990), and amino acid residues that are known to be important for receptor and effector protein interaction. This high sequence identity suggests that Gi class genes in *Fugu* and mammals may serve similar functions.

Gnaz is the most divergent Gi class gene. *Gnaz* is distinguished by the absence of the cysteine residue that is ADP ribosylated by pertussis toxin, and amino acid substitutions in the G1 box that probably contribute to a lower rate of GTP hydrolysis in Gαz than other Gi class gene products. The gene structure of *Gnaz* is also different from other Gi class genes, because it has an intron in the 5′ untranslated region and a single intron within the coding sequence, whereas other vertebrate Gi class genes have seven introns (Kaziro et al. 1991). The *Fugu* gene that we have identified is almost certainly the ortholog of mammalian *Gnaz* because these genes are 95% identical (human and rat are 98% identical to each other); they share several important amino acid sequences that contribute to its distinct functional characteristics (Fong et al. 1988; Matzuoka et al. 1988); the location of a single intron within the *Gnaz* coding sequence is conserved; and *Gnaz* is closely linked to the sodium–glucose cotransporter (SGLT1) in *Fugu* and humans (M.M. Sarwal and S. Brenner, unpubl.). A *Gnaz* homolog has not been found in invertebrates. Thus, we propose that *Gnaz* evolved from a vertebrate Gi class progenitor gene before the separation of marine and terrestrial vertebrate lineages and underwent rapid sequence divergence, possibly as an incompletely processed pseudogene intermediate (Wilkie et al. 1992), before becoming fixed in a newly acquired function.

Importantly, PCR amplification with degenerate primers did not identify orthologs of *Drosophila dgf*, snail Gαa, and *C. elegans gpa-1*, *gpa-2*, or *gpa-3*. We propose that these genes may have evolved in invertebrates to subserve specific functions in specialized cell types. The question remains whether vertebrates encode many more G protein alpha subunits not identified in previous

searches, including PCR amplification of cDNA (for review, see Wilkie et al. 1994), sequence analysis of expressed sequence tags [National Center for Biotechnology Information (NCBI) data base; Adams et al. 1995], or PCR amplification of *Fugu* genomic DNA. In addition to the 14 paralogs of mammalian Gα genes, *Fugu* contains a *Gnai2-like* and a *Gnas-like* gene, which may be pseudogenes, and *Gnap1*, which is highly conserved in three species of puffer fish and predominantly expressed in gills and spleen (data not shown). *Gnap1* is most closely related to the Gq class alpha subunits, and could be the ortholog of the mammalian *Gna15* gene. *Gnag*, the transducin homolog specifically expressed in taste cells (McLaughlin et al. 1992), is the only other mammalian Gα gene that we clearly did not identify in our screens. The fact that *Fugu* encodes most, if not all, of the G protein alpha subunits identified in mammals strongly supports *Fugu* as a model organism for vertebrate genome research.

METHODS

PCR

Amplification of genomic DNA (100 ng) was conducted in a reaction volume of 20 µl with 200 ng of each primer, 200 µM dNTPs, *Taq* buffer 6 [Stratagene; 10 mM Tris–HCl at pH 8.8, 75 mM KCl, 1.5 mM $MgCl_2$[, and 0.2 units of *Taq* polymerase. After an initial denaturing step at 94°C (5 min), DNA was amplified in 40 cycles: 94°C for 30 sec, 40–45°C for 45 sec, and 72°C for 2 min, and completed with a 5-min extension step at 72°C. Reaction products were separated on a 1% agarose gel (SeaKem GTG), stained with ethidium bromide, and the isolated DNA fragments were purified either with the Qiaex or Qiaquick DNA extraction kit (Qiagen). PCR fragments were digested and directionally cloned into *Bam*HI and *Eco*RI sites of KS+ (Stratagene) or blunt-end cloned into the EcoRV site of pBS II SK+. The resulting clones were screened and sequenced according to Wilkie et al. (1994).

PCR Primers

Degenerate pairs of sense and antisense primers were synthesized to complement the sequence encoding conserved amino acid motifs found in the G3, G4, and G5 boxes (see Fig. 1; Wilkie et al. 1994). Primers that were biased toward the Gi and Gq, Gs, or G12 classes were paired in separate PCR reactions (see Fig. 2). The DNA sequence (5′ to 3′) of the degenerate oligonucleotide sense primers [and the targeted amino acid motifs] are iMP19 [KWIHCF] CGGATCCAA(AG)TGGAT(I)CA(TC)TG(TC)TT; CT56 [(RK)W(MILF)(CWSR)CF] CGGATCCA(AG)(AG)TGG(TCA)T(I)(CG)A(AG)TG(TC)TT; CT48 [KWIQCF] GCGGATCCAA(AG)TGGAT(TCA)CA(AG)TG(TC)TT; and GP3F [ILFLNK(KQ)D] CT(CG)TTC(CT)TCAACAAG(AC)A(AG)GA. Antisense primers are TW8 [FLNK(RFL)D], GGAATTCTCIIG(TC)TT(AG)TTIA(AG)(AG)AA; CT49 [FLNKQD] CGGAATTC(AG)TC(TC)TG(TC)TT(AG)TT(ATCG)A(AG)(AG)AA; CT154 [TAIDTEN] GGGAATTCTT(CT)TCIGT(AG)TCIATIGCIGT; CT155 [VKDTIL(QH)] GGGAATTCTG(I)AG(I)AT(I)GT(AG)TC(TC)TT(I)AC; GP4R [TCATVDT] GTGTC(TG)GTGGCGCA(CG)GT(AGC)A(AT)GTG. Restriction endonuclease sites in the degenerate primers are underlined.

Isolation, Mapping, and Sequencing of Genomic Clones

A gridded *Fugu* genomic cosmid library with 4× genome coverage was screened with radiolabeled Gα PCR fragments from *Fugu* genomic DNA or mammalian Gα cDNAs. Filters were hybridized (Church and Gilbert 1984) at 55°C, rinsed at room temperature, then washed at 50°C for 5 min, 55°C for 5 min, and 60°C for 2 min, to final counts of 5–10 cpm and exposed to film for 48 hr. Duplicate positive cosmids were mapped by restriction enzyme digestion and Southern blot and hybridized with the probes that were originally used to screen the library. Cosmid restriction fragments that hybridized to the probes were sonicated to generate smaller fragments that were subcloned into the *Eco*RV site of pBS II SK+. DNA sequence was obtained (Sanger et al. 1977) from plasmid or internal oligonucleotide primers using Sequenase (U.S. Biochemical) or *Taq* polymerase (Perkin/Elmer) in an ABI 373A automated sequencer (Applied Biosystems). Templates for ABI sequencing were prepared by the method of Rosenthal et al. (1992, 1993). Coding regions were sequenced on both strands. Sequencing gels were read manually using DNAParrot, and sequence alignment, comparisons, and translation was done using DNAid and DNAStar. Comparisons with current releases of GenBank, SWISS-PROT, and European Molecular Biology Laboratory (EMBL) data bases were done using BLAST (Altschul et al. 1990). Sequence similarity was assessed using CLUSTAL. Transcription factor binding sites and promoter consensus elements were searched using Signalscan.

ACKNOWLEDGMENTS

We thank Kim Sweet and Tish Colwell for technical assistance, Juanita Satterfield for assistance in preparing the manuscript, and our colleagues for helpful comments. This work was supported in part by a Basil O'Connor Starter Scholar Award and National Institutes of Health grant DK47890 to TMW.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Adams, M.D., A.R. Kerlavage, R.D. Fleischman, R.A. Fuldner, C.J. Bult, N.H. Lee, E.F. Kirkness, K.G. Weinstock, J.D. Gocayne, O. White, et al. 1995. Initial assessment of human gene diversity and expression

patterns based upon 83 million nucleotides of cDNA sequence. *Nature (Suppl).* **377:** 3–174.

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Baxendale, S., S. Abdulla, G. Elgar, D. Buck, M. Berks, G. Micklem, R. Durbin, G. Bates, S. Brenner, S. Beck, and H. Lehrach. 1995. Comparative sequence analysis of the human and pufferfish Huntington's disease genes. *Nature Genet.* **10:** 67–76.

Birnbaumer, L. 1990. G proteins in signal transduction. *Annu. Rev. Pharmacol. Toxicol.* **30:** 675–705.

Bray, P., A. Carter, C. Simons, V. Guo, C. Puckett, J. Kamholz, A. Spiegel, and M. Nirenberg. 1986. Human cDNA clones for four species of Gαs signal transduction protein. *Proc. Natl. Acad. Sci.* **83:** 8893–8897.

Brenner, S., G. Elgar, R. Sandford, A. Macrae, B. Venkatesh, and S. Aparicio. 1993. Characterization of the pufferfish (*Fugu*) genome as a compact model vertebrate genome. *Nature* **366:** 265–268.

Buhl, A.M., N.L. Johnson, N. Dhanasekaran, and G.L. Johnson. 1995. G alpha 12 and G alpha 13 stimulate Rho-dependent stress fiber formation and focal adhesion assembly. *J. Biol. Chem.* **270:** 24631–24634.

Church, G.M. and W. Gilbert. 1984. Genomic sequencing. *Proc. Natl. Acad. Sci.* **81:** 1991–1995.

Davignon, I., M. Barnard, O. Gavrilova, K. Sweet, and T.M. Wilkie. 1996. Gene structure of murine *Gna11* and *Gna15*: Tandemly duplicated Gq class G protein alpha subunit genes. *Genomics* **31:** 359–366.

Elgar, G., F. Rattray, J. Greystrong, and S. Brenner. 1995. Genomic structure and nucleotide sequence of the p55 gene of the puffer fish *Fugu rubripes*. *Genomics* **27:** 442–446.

Fong, H.K.W., K.K. Yoshimoto, P. Eversole-Cire, and M.I. Simon. 1988. Identification of a GTP-binding protein alpha subunit that lacks an apparent ADP-ribosylation site for pertussis toxin. *Proc. Natl. Acad. Sci.* **85:** 3066–3070.

Gilman, A.G. 1987. G proteins: Transducers of receptor-generated signals. *Annu. Rev. Biochem.* **56:** 615–649.

Kaziro, Y., H. Itoh, T. Kozasa, M. Nakafuku, and T. Satoh. 1991. Structure and function of signal-transducing GTP-binding proteins. *Annu. Rev. Biochem.* **60:** 349–400.

Knol, J.C., A.R. van der Silk, E.R.L. van Kesteren, R.J. Plants, H. van Heerikhuizen, and E. Vreugdenhill. 1995. A novel G protein alpha subunit containing atypical guanine nucleotide-binding domains is differentially expressed in a molluscan nervous system. *J. Biol. Chem.* **270:** 18804–18808.

Kozasa, T., H. Itoh, T. Tsukamoto, and Y. Kaziro. 1988. Isolation and characterization of the human Gαs gene. *Proc. Natl. Acad. Sci.* **85:** 2081–2085.

Lochrie, M.A., J.E. Mendel, P.W. Sternberg, and M.I. Simon. 1991. Homologous and unique G protein alpha subunits in the nematode *Caenorhabditis elegans*. *Cell Reg.* **2:** 135–154.

McLaughlin, S.K., P.J. McKinnon, and R.F. Margolskee. 1992. Gustducin is a taste-cell-specific G protein closely related to the transducins. *Nature* **357:** 563–569.

Macrae, A.D. and S. Brenner. 1995. Analysis of the dopamine receptor family in the compact genome of the puffer fish *Fugu rubripes*. *Genomics* **25:** 436–446.

Mason, P.J., D.J. Stevens, L. Luzzatto, S. Brenner, and S. Aparicio. 1995. Genomic structure and sequence of the *Fugu rubripes* glucose-6-phosphate dehydrogenase gene (G6PD). *Genomics* **26:** 587–591.

Matsuoka, M., H. Itoh, T. Kozasa, and Y. Kaziro. 1988. Sequence analysis of cDNA and genomic DNA for a putative pertussis toxin-insensitive guanine nucleotide-binding regulatory protein alpha subunit. *Proc. Natl. Acad. Sci.* **85:** 5384–5388.

Mumby, S.M., R.A. Kahn, D.R. Manning and A.G. Gilman. 1990. G-protein α-subunit expression, myristoylation, and membrane association in COS cells. *Proc. Natl. Acad. Sci.* **87:** 265–269.

Powers, D.A. 1991. Evolutionary genetics of fish. *Adv. Genet.* **29:** 119–228.

Quan, F., W.J. Wolfgang, and M.A. Forte. 1989. The *Drosophila* gene coding for the a subunit of a stimulatory G protein is preferentially expressed in the nervous system. *Proc. Natl. Acad. Sci.* **86:** 4321–4325.

Rosenthal, A.R. and D.S. Charnock-Jones. 1992. New protocols for DNA sequencing with dye terminators. *DNA Sequence* **3:** 61–64.

Rosenthal, A., O. Coutelle, and M. Craxton. 1993. Large-scale production of DNA sequencing templates by microtitre format PCR. *Nucleic Acids Res.* **21:** 173–174.

Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad Sci.* **74:** 5463–5467

Sekido, Y., S. Badder, F. Latif, J-Y. Chen, F-M. Duh, M-H. Wei, J.P. Albanesi, C-C. Lee, M.I. Lerman, and J.D. Minna. 1996. Human semaphorins A(V) and IV reside in the 3p21.3 small cell lung cancer deletion region and demonstrate distinct expression patterns. *Proc. Natl. Acad. Sci.* **93:** 4120–4125.

Simon, M.I., M.P. Strathmann, and N. Gautam. 1991.

Diversity of G proteins in signal transduction. *Science* **252:** 802–808.

Talluri, S., A. Bhatt, and D.P. Smith. 1995. Identification of a *Drosophila* G protein alpha subunit (dGq alpha-3) expressed in chemosensory cells and central neurons. *Proc. Natl. Acad. Sci.* **92:** 11475–11479.

Wilkie, T.M. and S. Yokoyama. 1994. Evolution of the G protein alpha subunit multigene family. In *Molecular evolution of physiological processes* (ed. D.M. Fambrough), pp. 249–270. Rockefeller University Press, New York, NY.

Wilkie, T.M., D.J. Gilbert, A.S. Olsen, X.N. Chen, T.T. Amatruda, J.R. Korenberg, B.J. Trask, P. deJong, R.R. Reed, M.I. Simon, N.A. Jenkins, and N.G. Copeland. 1992. Evolution of the mammalian G protein α subunit multigene family. *Nature Genet.* **1:** 85–91.

Wilkie, T.M., A.M. Aragay, A.J. Watson, and M.I. Simon. 1994. Design of degenerate oligonucleotide primers for the cloning of G protein α subunits. *Methods Enzymol.* **237:** 327–344.

Received July 16, 1996; accepted in revised form September 18, 1996.