Genomic Structure, Evolution, and Expression of Human *FLII,* a Gelsolin and Leucine-Rich-Repeat Family Member: Overlap with *LLGL*

Hugh D. Campbell,[1] Shelley Fountain, Ian G. Young,* Charles Claudianos,† Jörg D. Hoheisel,‡ Ken-Shiung Chen,§ and James R. Lupski§

*Molecular Evolution and Systematics Group and Centre for Molecular Structure and Function, Research School of Biological Sciences, The Australian National University, Canberra, ACT 2601, Australia; *Division of Biochemistry and Molecular Biology, John Curtin School of Medical Research, The Australian National University, Canberra, ACT 2601, Australia; †Division of Botany and Zoology, The Australian National University, and Division of Entomology, CSIRO, Canberra, ACT 2601, Australia; ‡Molecular-Genetic Genome Analysis, German Cancer Research Center, Im Neuenheimer Feld 506, D-69120 Heidelberg, Germany; and §Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030-3498*

Received December 4, 1996; accepted March 6, 1997

The *Drosophila melanogaster flightless-I* gene is involved in cellularization processes in early embryogenesis and in the structural organization of indirect flight muscle. The encoded protein contains a gelsolin-like actin binding domain and an N-terminal leucine-rich repeat protein–protein interaction domain. The homologous human *FLII* gene encodes a 1269-residue protein with 58% amino acid sequence identity and is deleted in Smith–Magenis syndrome. We have cloned the *FLII* gene and determined its nucleotide sequence (14.1 kb). *FLII* has 29 introns, compared with 13 in *Caenorhabditis elegans* and 3 in *D. melanogaster*. The positions of several introns are conserved in *FLII*-related genes and in the domains and subdomains of the gelsolin-like regions giving indications of gelsolin gene family evolution. In keeping with its function in indirect flight muscle in *Drosophila,* the human *FLII* gene was most highly expressed in muscle. The *FLII* gene lies adjacent to *LLGL,* the human homologue of the *D. melanogaster* tumor suppressor gene *lethal(2) giant larvae.* The 3′ end of the *FLII* transcript overlaps the 3′ end of the *LLGL* transcript, and the corresponding mouse genes *Fliih* and *Llglh* also overlap. The overlap region contains poly(A) signals for both genes and is strongly conserved between human and mouse. © 1997 Academic Press

INTRODUCTION

Depending on their severity, mutations in the *Drosophila melanogaster flightless-I* (*fliI*) gene (Homyk and Sheppard, 1977) cause flightlessness with abnormal myofibrillar arrangements in the indirect flight muscles (Deak *et al.*, 1982; Miklos and de Couet, 1990) or lethality with abnormal gastrulation and only partial cellularization of the syncytial blastoderm (Perrimon *et al.*, 1989). We have recently characterized the cDNA for *Drosophila fliI* (Campbell *et al.*, 1993) and shown that the gene encodes a 1256-amino-acid protein with a gelsolin-like domain characteristic of proteins that interact with actin to cap and/or sever actin filaments (Hartwig and Kwiatkowski, 1991). The fliI protein also carries an N-terminal protein–protein interaction domain consisting of leucine-rich repeats (LRRs) (Kobe and Deisenhofer, 1995).

Highly conserved *fliI* homologues are also present in *Caenorhabditis elegans* and humans (49 and 58% amino acid sequence identity, respectively) (Campbell *et al.,* 1993; Claudianos and Campbell, 1995), suggesting conservation of the function(s) of this protein over an evolutionary period of at least 500 million years. Interestingly, the human *FLII* gene, which maps to human chromosome 17p11.2, is deleted in Smith–Magenis syndrome (SMS), a relatively common (1 in 25,000 live births) microdeletion syndrome involving developmental abnormalities and mental retardation (Chen *et al.*, 1995).

We report here the isolation of genomic cosmid clones spanning *FLII,* the nucleotide sequence of the gene, and the demonstration that it is most highly expressed in muscle. Conservation of intron positions within subunit domains and across other members of the gelsolin gene family gives some indications of the evolution of these genes, and this was also examined by phylogenetic analysis of the protein domains. The *LLGL* gene (Strand *et al.*, 1995; Koyama *et al.*, 1996), the human homologue of the *D. melanogaster* tumor suppressor gene *lethal(2) giant larvae (l(2)gl),* is shown to lie adja-

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession No. U80184.

[1] To whom correspondence should be addressed at Molecular Evolution and Systematics Group and Centre for Molecular Structure and Function, RSBS, The Australian National University, GPO Box 475, Canberra, ACT 2601, Australia. Telephone: 61-6-2495080. Fax: 61-6-2494437. E-mail: Hugh.Campbell@anu.edu.au.

0888-7543/97 $25.00

cent to the *FLII* gene in the opposite transcriptional orientation. The 3′ ends of the transcripts overlap for both the human *FLII* and *LLGL* genes and the corresponding murine genes *Fliih* and *Llglh*. The overlap region contains poly(A) signals for both genes and is highly conserved between human and mouse.

MATERIALS AND METHODS

Cloning of the FLII gene. Cosmids were isolated from a gridded chromosome 17 library in SuperCos I (Stratagene) from Los Alamos National Laboratory (LA17NC01). The library was prepared from chromosomes flow-sorted from the mouse–human hybrid cell line 38L-27 (Kallioniemi *et al.*, 1994). Cosmids were also isolated from the ICRF gridded chromosome 17 Reference Library (Zehetner and Lehrach, 1994). Screening was done with the 4.1-kb human *FLII* cDNA (Campbell *et al.*, 1993) and cDNA probes from the 5′ (274-bp *Eco*RI) and 3′ (378-bp *Not*I) ends. DNA probes were labeled by random primer incorporation of [α-^{32}P]dCTP. Cosmid DNA was isolated using Qiagen kits. *Not*I fragments were subcloned into pBluescript KS(+). Standard recombinant DNA methods were as described (Sambrook *et al.*, 1989).

Sequence analysis. End sequencing was done using Applied Biosystems PRISM dye primer and dye terminator reagents. Sonicated fragments of the 13.7-kb *Not*I fragment were cloned into M13mp10 (Deininger, 1983) and sequenced with dye primer reagents. Sequences were determined on an Applied Biosystems 373A or 377 DNA sequencer. Specific primers were used to ensure coverage of both strands. The only exception was part of the final exon and a few bases 3′ to it that were determined on one strand only, using both dye primers and dye terminators. The sequence of the final exon agreed exactly with the cDNA sequence (Campbell *et al.*, 1993). The overall sequencing redundancy level was >eightfold.

Computer methods. Sequence assembly and analysis used Staden (1987), Genetics Computer Group (Devereux *et al.*, 1984), MacVector (Kodak), and EditView (Applied Biosystems) software. Sequence databases were searched using BLAST options at NCBI, Bethesda. Sequence alignments of proteins and their domains were obtained using the GCG programs Gap and PileUp, with gap weight 3.0 and gap length weight 0.1. End gaps were weighted like other gaps. Protein distance matrix analysis was performed with the Phylogenetic Inference Package, PHYLIP, Version 3.5c (Felsenstein, 1993). One hundred bootstrap replications were carried out via the program SEQBOOT. PRODIST, in conjunction with the PAM001 matrix, produced distance datasets from which a neighbor-joining tree was calculated using NEIGHBOR (Saitou and Nei, 1987). Maximum parsimony analysis was carried out using PAUP (Swofford, 1993). Unrooted protein trees were rooted to severin as the ancestral taxon using the outgroup method.

Southern and Northern blot analysis. Human placental DNA (10 μg) was digested, separated on a 0.8% agarose gel, and blotted onto reinforced nitrocellulose membrane (Schleicher & Schuell, BA-S 85). The membrane was hybridized with the ^{32}P-labeled 4.1-kb *FLII* cDNA fragment (Campbell *et al.*, 1993). The final wash was at 65°C in 1× or 0.1× SSC, 0.1% SDS. A human MTN blot, Clontech Catalog No. 7760-1, was hybridized to the *FLII* cDNA probe or a human β-actin probe (Clontech) at 65°C in 5× SSC, 50 m*M* sodium phosphate buffer, pH 6.8, 10× Denhardt's solution, 2% SDS, 10 m*M* EDTA, 1 m*M* sodium pyrophosphate, 1 m*M* ATP, 100 μg/ml sonicated, denatured salmon sperm DNA and washed at 65°C in 1× SSC, 0.1% SDS. Autoradiography was performed with intensifying screens at −80°C. Signal intensity was quantified with a Molecular Dynamics PhosphorImager.

PCR. For nested PCR, 10 pmol of the outer primer 5′-AAG AGG ACA GGT GGG GGC CCA GCA C-3′ (*LLGL* cDNA, GenBank D50550, nucleotides 3487–3511) was used together with 10 pmol of an M13 reverse primer (5′-AGC GGA TAA CAA TTT CAC ACA GGA-3′) in a 50-μl reaction with 1 μl boiled human hippocampus cDNA library (Stratagene Catalog No. 936205) as template in a Perkin–Elmer system 9600 PCR machine using AmpliTaq DNA polymerase. After 1.5 min at 95°C, 40 cycles of 95°C, 30 s; 55°C, 30 s; and 72°C, 60 s were carried out, followed by 4.5 min at 72°C. An aliquot (1 μl) of this reaction was then used as template in a second PCR with the inner primer 5′-CAT CCT TCC CCC TCA CTT TGC AGA G-3′ (*LLGL* cDNA, GenBank D50550, nucleotides 3539–3563) and a T3 primer (5′ATT AAC CCT CAC TAA AGG GA-3′), under the same reaction conditions. The 600-bp product was phosphorylated, purified on an agarose gel, and cloned into M13mp10 for dye primer sequencing.

RESULTS

Cloning and Analysis of FLII Genomic DNA

We previously reported the isolation of three cosmids spanning the 5′ and 3′ ends of the *FLII* gene (Chen *et al.*, 1995). We have isolated further cosmids hybridizing to *FLII* cDNA probes from chromosome 17 cosmid libraries (Kallioniemi *et al.*, 1994; Zehetner and Lehrach, 1994). A restriction map (*Eco*RI and *Not*I) of the cloned genomic DNA in 11 cosmids was generated, and hybridization to the *FLII* cDNA and probes from the 5′ and 3′ ends of the cDNA localized the *FLII* gene (not shown). Southern blot analysis of human genomic DNA (Fig. 1A) with the *FLII* cDNA probe showed that the gene was present as a single copy and was consistent with the map of the cloned *FLII* region and with the nucleotide sequence of the gene (see below), indicating the absence of major rearrangements in the cloned DNA. The same pattern of bands was observed on the genomic Southern blot at final wash stringencies of 1× or 0.1× SSC at 65°C, and with exposure times up to 68 h, indicating that there are no other closely related genes present detectable under these conditions.

Northern Analysis of FLII Gene Expression

A Northern blot containing 2 μg per lane of poly(A)$^+$ RNA from human heart, brain, placenta, lung, liver, skeletal muscle, kidney, and pancreas was hybridized to the *FLII* cDNA probe. A band of ~4.4 kb was visible in all lanes after 2.4 h autoradiography (Fig. 1B). A strong band of this size was observed in all lanes after 19 h autoradiography, with no additional bands visible (not shown). After background subtraction, the signal intensities, relative to liver set as 1.0, were heart, 2.0; brain, 1.0; placenta, 0.8; lung, 1.5; liver, 1.0; skeletal muscle, 7.7; kidney, 0.8; pancreas, 0.8. Thus, *FLII* is expressed in all these tissues, with strongest expression in skeletal muscle, followed by heart, and then lung. Hybridization to a human β-actin probe (Fig. 1C) confirmed approximately equal loading of poly(A)$^+$ RNA in each lane.

Sequence Analysis of the FLII Gene

Cosmid c5C2 contained a 13.7-kb *Not*I fragment that hybridized to the 5′ end cDNA probe and extended to the *Not*I site present near the 3′ end of the cDNA at nucleotide 3709 (Campbell *et al.*, 1993). The 3′ end of the gene was present on an adjacent 9-kb *Not*I frag-

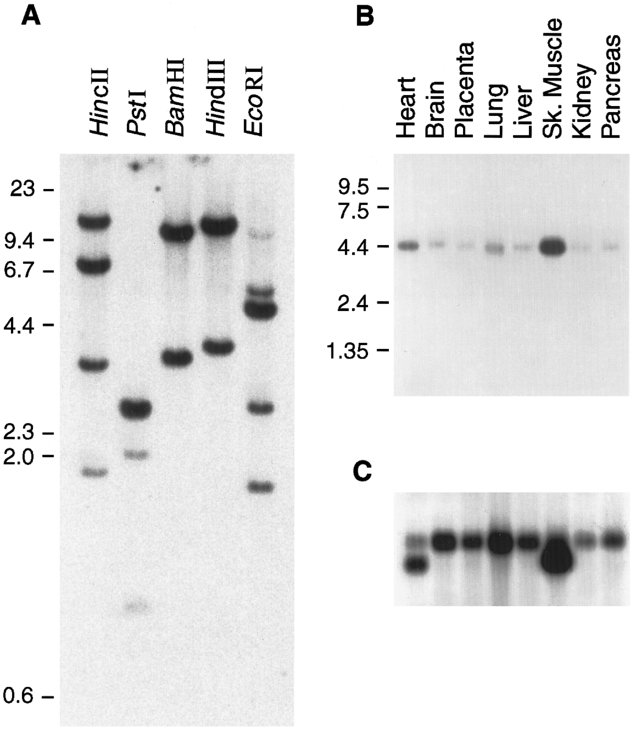

FIG. 1. Southern blot analysis of human genomic DNA with the *FLII* cDNA probe and Northern blot analysis of *FLII* gene expression in various tissues. **(A)** High-molecular-weight human placental DNA (10 μg) was digested with restriction endonucleases *Eco*RI, *Hin*dIII, *Bam*HI, *Pst*I and *Hin*cII, separated on a 0.8% agarose gel, transferred to reinforced nitrocellulose membrane, and hybridized to the 4.1-kb *FLII* cDNA probe. The blot was exposed to X-ray film for 19.5 h at −80°C. **(B)** A Northern blot containing human poly(A)$^+$ RNA (2 μg) from heart, brain, placenta, lung, liver, skeletal muscle (sk. muscle), kidney, and pancreas was hybridized to the 4-kb *FLII* cDNA probe. The blot was exposed to X-ray film for 2.4 h at −80°C. **(C)** Northern blot from **B** stripped and hybridized to a human β-actin probe. In all cases, the final wash conditions were 1× SSC, 0.1% SDS at 65°C. The sizes of markers in kilobases are indicated to the left of the blots.

ment. The complete nucleotide sequence of the 13.7-kb *Not*I fragment was determined on both strands by automated sequencing. One end of the 9-kb *Not*I fragment matched the 3′ end of the cDNA exactly, extending from the *Not*I site at 3709 in the cDNA sequence to the poly(A) attachment site (Campbell *et al.*, 1993), with no introns present. The nucleotide sequence of the *FLII* gene that we have determined covers 14131 bp.

An EST clone, GenBank Accession No. R33910, extending 5′ to our previously determined cDNA sequence (Campbell *et al.*, 1993), was identified and resequenced, extending the cDNA sequence a further 37 bp upstream across the putative ATG translation initiation codon. This is the first ATG in the open reading frame after an in-frame TAA stop codon upstream in the genomic sequence. The previous cDNA sequence was missing only the AT of the ATG translation initiation codon. Although we have not determined the exact position at which transcription of *FLII* begins, the results of Northern analysis suggest that the 5′ end of R33910 is probably not far from this point. Examination of the 207 bp of genomic sequence extending upstream from the 5′ end of R33910 reveals a very GC-rich region in which a TTAAA motif is embedded, possibly representing the TATA box for *FLII*. The TTAAA motif is located 61 bp upstream from the 5′ end of R33910. The 144 bp of the 5′ sequence extending from within the potential TATA box to the *Sau*3AI cloning site exactly matches the sequence of a CpG island clone (Cross *et al.*, 1994; GenBank Accession No. Z59965), indicating that this is likely to represent at least a portion of the 5′ regulatory region of *FLII*.

Exon/Intron Structure of the FLII Gene

A schematic view of the human *FLII* gene is presented in Fig. 2, and a summary of the exon/intron boundaries is presented in Fig. 3. The *FLII* gene contains 29 introns, compared with 3 in the *D. melanogaster fliI* gene (GenBank Accession Nos. U01182 and U28044) (Campbell *et al.*, 1993; de Couet *et al.*, 1995) and 13 in the *C. elegans* gene (GenBank Accession Nos. U01183, M77697, L18807, and L07143) (Campbell *et al.*, 1993; Sulston *et al.*, 1992). The conservation of intron positions between the different characterized members of the *FLII* gene family (Fig. 4A) provides further strong support for the common evolutionary origin of these genes and indicates that a number of the present-day *FLII* gene introns were present in the *FLII* gene homologue in the latest common ancestor for humans, *D. melanogaster*, and *C. elegans* at least 500 million years ago.

Analysis of Exon/Intron Structure of Gelsolin Family Members

Apart from the *FLII* genes, the only other members of the gelsolin gene family for which detailed information is available on exon/intron structure are human villin (Pringault *et al.*, 1991), human Cap G (Mishra *et al.*, 1994), *Dictyostelium discoideum* protovillin (Hoffman *et al.*, 1993), and human gelsolin (partial; Kwiatkowski *et al.*, 1988; Witke *et al.*, 1995). Recently, a novel member of the family has been found in *C. elegans* (GenBank Accession No. Z70755, reading frame K06A4.3) (Sulston *et al.*, 1992).

The human villin gene (Pringault *et al.*, 1991) contains 18 introns and spans 25 kb. A number of introns are conserved in position between human *FLII* and villin, and one of these corresponds to the single intron of *D. discoideum* protovillin (Hoffman *et al.*, 1993) (Fig. 4). The *C. elegans* protein K06A4.3 (GenBank Accession No. Z70755) (Sulston *et al.*, 1992) consists of domain 1 (subdomains 1, 2, and 3) followed by subdomain 6 (Fig. 4), and 5 of its 9 introns correspond in position with introns in human villin (Fig. 4A).

The human *CAPG* gene (Mishra *et al.*, 1994) contains 9 introns and spans 16.6 kb. Intron 18 in domain 1 of the gelsolin-like portion of *FLII* is in exactly the same

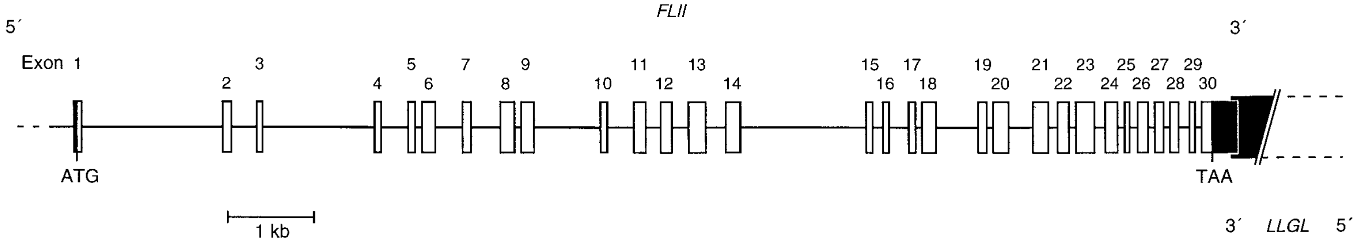

FIG. 2. Schematic view of the structure of the human *FLII* gene. The exons of the gene are boxed and numbered 1–30. Shaded portions of the boxes represent untranslated regions, and the open boxes represent coding sequence. The ATG initiation codon and TAA termination codon are marked. The 3′ end of the *LLGL* gene (Strand *et al.*, 1995; Koyama *et al.*, 1996) is also shown overlapping with the 3′ end of *FLII.* The scale is indicated.

position as intron 7 of *CAPG* (Fig. 4A). No other introns appear to be conserved in position between *CAPG* and *FLII.* None of the introns present near the 5′ end of the 70-kb human gelsolin gene (Kwiatkowski *et al.*, 1988; Witke *et al.*, 1995) are conserved in *FLII.* Mishra *et al.* (1994) state that "The first 5 exons of *CAPG* and gelsolin reveal nearly identical intron–exon junctions at 3 of 5 sites" based on unpublished information on the human and mouse gelsolin gene structures. Overall, the data indicate that Cap G is more closely related to domain 1 of gelsolin family members than to domain 2. The data also indicate that some present-day introns occurred in the latest common ancestor of gelsolin family members, prior to the divergence of lines leading to humans and *Dictyostelium.*

56	TTCAAG	**GT**GAGCCG	- intron 1, 1608 bp -	TCCCCC**AG**	GGCGGC	67
167	AAGCTG	**GT**AAGGGG	- intron 2, 311 bp -	CGGCCC**AG**	GAACAC	178
239	CTGCGC	**GT**GAGTGC	- intron 3, 1262 bp -	ATTCCC**AG**	GCCATC	250
320	GTCCTG	**GT**CAGTGG	- intron 4, 328 bp -	CTACCC**AG**	GACTTG	331
406	CAACAG	**GT**GCCAGG	- intron 5, 86 bp -	ACCCGT**AG**	CATCGA	417
568	GCTCCG	**GT**GGGCGC	- intron 6, 338 bp -	CCTGGC**AG**	GCAGCT	579
672	TCGCAG	**GT**CAGGCA	- intron 7, 359 bp -	CACCCC**AG**	ACGTGG	683
848	CTGCCC	**GT**GCGTCT	- intron 8, 85 bp -	CCCTCC**AG**	TCAGCC	859
1006	CTGCAG	**GT**GCTGGG	- intron 9, 744 bp -	CTTCAC**AG**	GTGCCC	1017
1091	ATCGAG	**GT**CAGGCA	- intron 10, 325 bp -	TGGGGC**AG**	GTCCTG	1102
1239	CAGCTG	**GT**GAGTGG	- intron 11, 182 bp -	CCCCTA**AG**	CAGGGA	1250
1376	CAGGAG	**GT**GAGCCA	- intron 12, 202 bp -	GCCCCC**AG**	GAGAGC	1387
1589	CTCAAG	**GT**GAGGGC	- intron 13, 247 bp -	GCTTCC**AG**	ACCTTT	1600
1769	CTGCAG	**GT**GCCAGC	- intron 14, 1429 bp -	ACACAC**AG**	GTGTTT	1780
1852	CACCAG	**GT**GAGGGT	- intron 15, 124 bp -	GCGCCC**AG**	GATGTA	1863
1927	CCCAAG	**GT**GAGCCC	- intron 16, 243 bp -	TTGTCC**AG**	GTTTGT	1938
2011	GGCCAG	**GT**ACAAGG	- intron 17, 78 bp -	GCTCTC**AG**	GCTCTT	2022
2183	TACAAG	**GT**GAGCCC	- intron 18, 516 bp -	CCTGCC**AG**	GTGGGC	2194
2288	CGGCTG	**GT**AAGAGA	- intron 19, 80 bp -	GTCCGC**AG**	CTGCAG	2299
2480	GCGCAG	**GT**GCGCTT	- intron 20, 299 bp -	CCTGGC**AG**	GTGTTC	2491
2669	GCCGAG	**GT**GGGGGC	- intron 21, 116 bp -	ACCTGC**AG**	GCGGAG	2680
2809	CTGCAG	**GT**ACCACC	- intron 22, 84 bp -	CCCGGC**AG**	GTACTG	2820
3044	CTGGAG	**GT**GTGCCT	- intron 23, 131 bp -	ATCCCC**AG**	GTGGTA	3055
3199	CACCCG	**GT**GCCTGG	- intron 24, 85 bp -	TTCCAC**AG**	GTGCAT	3210
3260	CTCAAG	**GT**GGGGTT	- intron 25, 99 bp -	CTCCCC**AG**	GTTCCC	3271
3389	AAGCAG	**GT**CAGGAG	- intron 26, 83 bp -	TCCCGC**AG**	GTTATC	3400
3496	CTTCCG	**GT**GAGGCC	- intron 27, 83 bp -	ATTCCC**AG**	GTGCTC	3507
3602	CAAGAG	**GT**GTGATG	- intron 28, 135 bp -	CCCCAC**AG**	GTCTAC	3613
3668	TGCCAG	**GT**AATCTG	- intron 29, 81 bp -	CTTGCC**AG**	GTATAT	3679

FIG. 3. The exon–intron boundaries of the human *FLII* gene. The numbers indicate the positions of the 5′ and 3′ nucleotides within the human *FLII* cDNA sequence (GenBank Accession No. U01184) (Campbell *et al.*, 1993). On each line, the final six nucleotides of each exon are given, followed by the first eight bases of the intron. The intron number is given, with its total size, then the last eight bases of the intron, followed by the first six bases of the next exon. The GT and AG at the 5′ and 3′ ends of the introns are in boldface type and underlined.

Analysis of Domains 1 and 2 of the Gelsolin-like Portion of Gelsolin Family Members

Gelsolins, adseverins, villins, and FLII proteins show evidence of an internal duplication, containing two copies of a unit similar to severin, fragmin, and Cap G (Campbell *et al.*, 1993; Claudianos and Campbell, 1995; Way and Weeds, 1988; Bazari *et al.*, 1988). We performed multiple alignments on the monomeric family members together with the separated domains 1 and 2 of dimeric family members and used these alignments to examine aspects of the evolution of the family. First, we examined the position of introns within the domains to see whether there is evidence of conservation of introns from the precursor from which domains 1 and 2 arose by gene duplication, and we found several cases where introns are in similar positions in domain 1 vs domain 2 of the gelsolin-like portions of family members (Fig. 4B). Second, we performed phylogenetic analyses on the separated domains to try to clarify the sequence of events by which the various family members have arisen. A distance matrix tree is shown in Fig. 5. Maximum parsimony methods give similar results (not shown). Cap G and the separated domains 1 all segregate together, and similarly, the separated domains 2 all segregate together. There is little evidence for gene conversion between the portions of the genes encoding domains 1 and 2 having played a significant role during the evolution of family members.

Overall, the results strongly indicate that the Cap G protein is more closely related to domain 1 of the dimeric family members and support the hypothesis that *CAPG* arose by loss of domain 2 from a gelsolin-like precursor gene (Claudianos and Campbell, 1995). In further support of this, the pattern of insertions and deletions (indels) in Cap G is identical (with one exception) with that of segment 1 of the gelsolins, whereas significant differences are present compared with all other segments 1 and all segments 2 (not shown), mir-

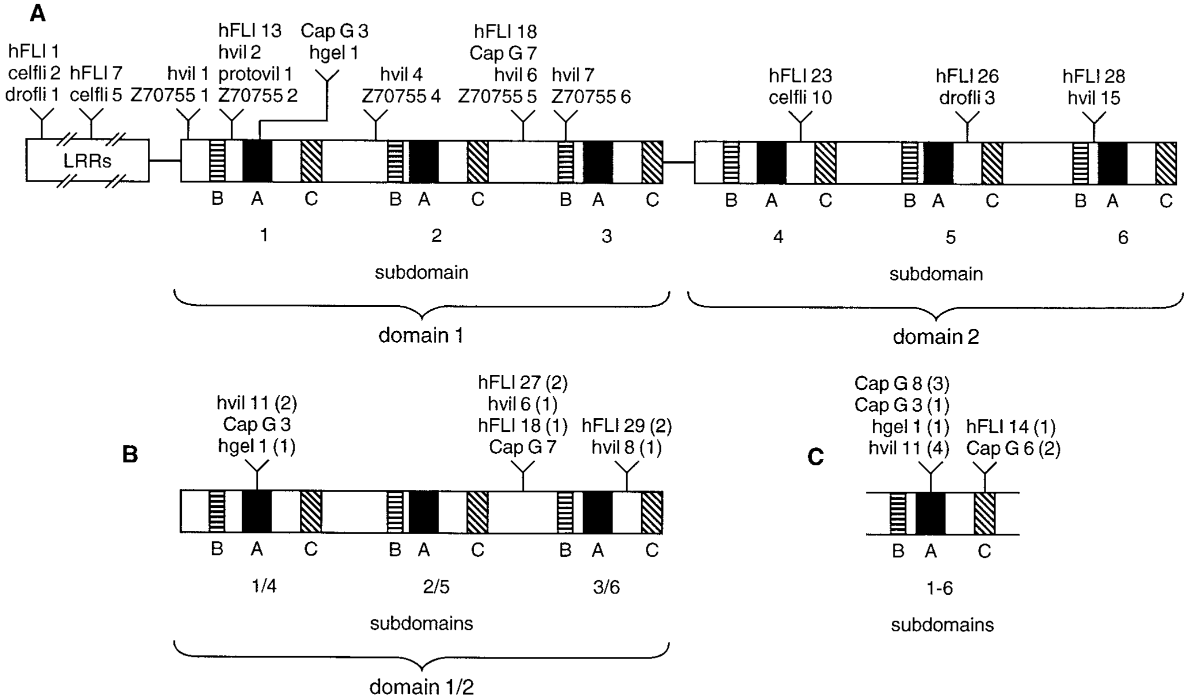

FIG. 4. Conservation of intron position. Domains 1 and 2, subdomains 1–6, and the B, A, and C motifs of the subdomains are indicated schematically. hFLI, human *FLII* gene; celfli, *C. elegans fliI* gene; drofli, *D. melanogaster fliI* gene; hvil, human villin gene; protovil, *D. discoideum* protovillin gene; Z70755, *C. elegans* gene encoding reading frame K06A4.3 in GenBank entry Z70755; Cap G, human Cap G gene; hgel, human gelsolin gene. The intron number follows the abbreviation for each gene. Introns conserved within 1 base are indicated. **(A)** Introns conserved between different members of the gelsolin gene family. **(B)** Introns conserved between domains 1 and 2 of family members. The numbers in parentheses indicate the domains (1 or 2) within which the introns occur. **(C)** Introns conserved between the subdomains of family members. The numbers in parentheses indicate the subdomains (1–6) within which these introns occur.

roring the phylogenetic trees (Fig. 5; Claudianos and Campbell, 1995). Domains 1 and 2 contain evidence of a triplication (Way and Weeds, 1988; Bazari *et al.*, 1988). In this work, we have identified some introns that may have been present in the ancestral unit prior to the triplication (Fig. 4C).

The FLII and LLGL Transcripts Overlap

The opposite end of the 9-kb *Not*I fragment from that containing the 3′ portion of *FLII* was sequenced, and a very strong match to the previously described *LLGL* gene was found (Strand *et al.*, 1995; Koyama *et al.*, 1996). After a match of 106 bp extending from the *Not*I site, a short intron of 74 bp is present, and then the match resumes (Fig. 6). The close proximity of *FLII* and *LLGL* was also demonstrated by the fact that the *FLII* cDNA probe and a PCR-generated probe for *LLGL* (Strand *et al.*, 1995; Koyama *et al.*, 1996) identified five of the identical cosmids (c5C2, c5F6, c92C10, c110H8, and c157D9) when used to screen a chromosome 17 cosmid library (Kallioniemi *et al.*, 1994).

Reported cDNAs for *LLGL* are truncated at the 3′ end (Strand *et al.*, 1995; Koyama *et al.*, 1996). Therefore we used nested PCR on human hippocampus cDNA to extend the 3′ end of the *LLGL* cDNA. For this purpose, primers were designed to avoid an *Alu* element present at the 3′ end of the available *LLGL* sequence. As expected, one end of the 600-bp PCR product matches *LLGL* cDNA (GenBank Accession No. D50550) to the end of the available sequence (99.6% identity over 268 bases). The other end overlaps the 5′ end sequence (98.9% identity over 277 bases) from a single human EST clone (GenBank Accession No. W35195) that has been sequenced from both ends. The 3′ end sequence of this clone (GenBank Accession No. W23681), which contains a poly(A) signal and tail (original sequence trace downloaded from http://genome.wustl.edu/est/esthmpg.html), therefore represents the 3′ end of *LLGL* cDNA and was found to be 100% identical to the 3′ end of the *FLII* gene sequence we have determined (Fig. 7), establishing that the *FLII* and *LLGL* genes overlap.

We are also sequencing genomic *Fliih*, the mouse homologue of *FLII*. The sequence of 2.5 kb of *Fliih* 3′ to the final exon shows extensive sequence identity to mouse *Llglh* on the opposite strand, allowing for several introns (unpublished results). The sequence of *Llglh* cDNA includes the complete 3′ end (Tomotsune *et al.*, 1993). The final exons of *Fliih* and *Llglh* overlap in the region of the poly(A) sites as illustrated in Fig. 7. Mouse brain *Fliih* cDNAs that we have analyzed are polyadenylated at the 5′ most of the poly(A) sites (Fig. 7), but human *FLII* EST cDNA clones from other tissues are also polyadenylated at the other site shown in Fig. 7, which also uses the variant ATTAAA poly(A)

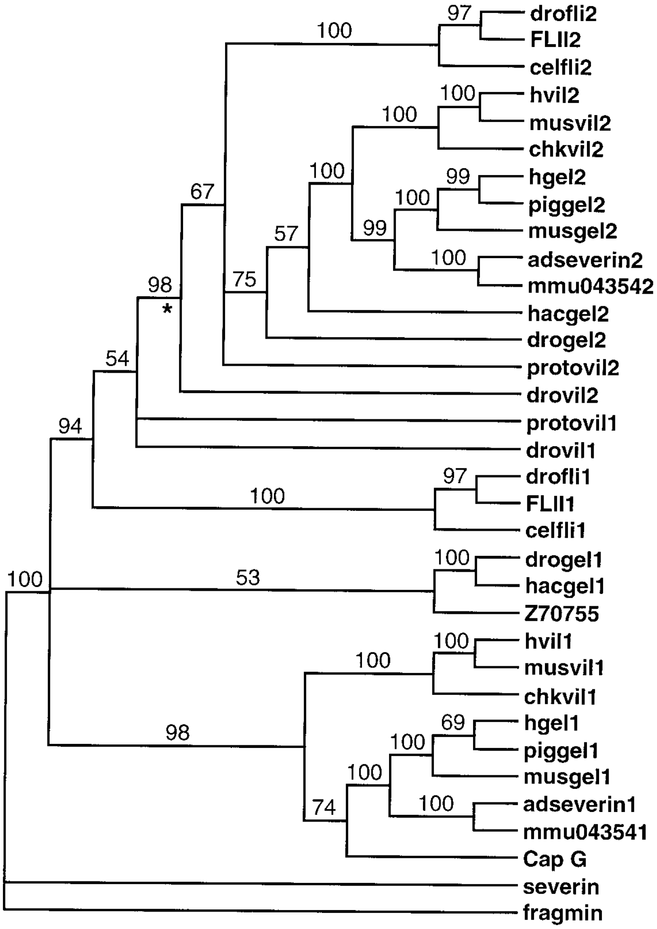

FIG. 5. Distance matrix analysis of the separated domains 1 and 2 of members of the gelsolin gene family. Abbreviations of the names of the gelsolin family members are as in this paper or as used previously (Claudianos and Campbell, 1995), with 1 or 2 appended, as appropriate, to indicate gelsolin-like domain 1 or 2. drovil, *D. melanogaster* quail protein (Mahajan-Miklos and Cooley, 1994). Bootstrap values are indicated at the nodes. The node from which the segments 1 and 2 segregate is indicated by an asterisk.

signal (unpublished results). Although we have only detected mouse brain *Fliih* cDNA clones polyadenylated at the 5′ most of the poly(A) sites (Fig. 7) (Campbell *et al.*, 1993), it seems likely that mouse *Fliih* transcripts polyadenylated using the additional poly(A) site as in *FLII* (Fig. 7) may also occur, possibly in a tissue-specific manner.

DISCUSSION

In the present study we have cloned and sequenced the chromosomal *FLII* gene, the human homologue of the *D. melanogaster flil* gene (Campbell *et al.*, 1993). The *FLII* gene spans 14 kb of genomic DNA and is smaller than the genes for other mammalian members of the gelsolin gene family such as human gelsolin (70 kb) (Kwiatkowski *et al.*, 1988), human villin (25 kb) (Pringault *et al.*, 1991), and the monomeric Cap G (16.6 kb) (Mishra *et al.*, 1994), although it contains many more introns than the latter two genes. The human gelsolin gene contains a minimum of 13 introns, but only the 5′ end has been characterized at the sequence level (Kwiatkowski *et al.*, 1988).

FLII and its *Drosophila* and *C. elegans* homologues encode an N-terminal LRR domain found in more than 60 proteins (Kobe and Deisenhofer, 1995). In general, this domain is involved in protein–protein interactions (Kobe and Deisenhofer, 1995). Recently the first 3D structure of an LRR protein, that of the porcine ribonuclease inhibitor and of its complex with ribonuclease, has been determined (Kobe and Deisenhofer, 1995). The ligand for the LRR domain of the FLII protein is unknown, but may be a member of the ras family of proteins, based on analysis of the close relationship of the LRR domain of FLII proteins with LRR domains known to interact with ras, including those of yeast adenylate cyclase and the mammalian Rsu-1 proteins (Claudianos and Campbell, 1995).

The FLII proteins also contain a C-terminal gelsolin-like domain (Campbell *et al.*, 1993). This internally duplicated domain is present in gelsolins, villins, adseverins, and the FLII proteins. A monomeric version of the domain is present in the slime mold proteins severin and fragmin and mammalian Cap G (Hartwig and Kwiatkowski, 1991; Mishra *et al.*, 1994). Several introns conserved in position between domains 1 and 2 may have been present in the ancestral gene prior to the duplication. The monomeric domains contain evidence of an internal triplication (Way and Weeds, 1988; Bazari *et al.*, 1988). Therefore, the evolution of the gelsolin-like domain has proceeded by way of a triplication, followed by a duplication (Way and Weeds, 1988; Bazari *et al.*, 1988; Claudianos and Campbell, 1995). Some introns occur at common positions within the A and C motifs of the triplicated subdomains, suggesting they may have been present prior to the triplication.

Although the concept that the monomeric Cap G protein represents an intermediate species on this evolutionary pathway is attractive, we suggested previously

```
                   .         .         .         .         .
LLGL      GCGGCCGCTCTATAAGCTGAGCACAGCTGGCCTCTTCCAGACAGACTGTG 1495
          ||||||||||||||||||||||||||||||||||||||||||||||||||
9kb NotI  GCGGCCGCTCTATAAGCTGAGCACAGCTGGCCTCTTCCAGACAGACTGTG 50

                   .         .         .         .         .
LLGL      AGCACGCTGACAGCCTGGCCCAGGCTGCCGAGGACGACTGGCCACCCTTC 1545
          |||||||||||||||||||||||||||||||||||:||||||||||||||
9kb NotI  AGCACGCTGACAGCCTGGCCCAGGCTGCCGAGGACNACTGGCCACCCTTC 100

                   .         .         .         .         .
LLGL      CGCAAG............................................ 1551
          ||||||
9kb NotI  CGCAAGGTGGGCCCCTCCCCTGGCCCTGATGAGCTGGTCCTGACCCCGCC 150

                   .         .         .         .         .
LLGL      ..............................GTGGGCTGCTTCGATCCCTA 1571
                                        ||||||||||||||||||||
9kb NotI  TGGACTCATTACTCCTTCCCATCCCTCTAGGTGGGCTGCTTCGATCCCTA 200
```

FIG. 6. Mapping of *LLGL* near *FLII*. The nucleotide sequence of the first 200 bp of one end of the human genomic 9-kb *Not*I fragment from cosmid c5C2 is shown aligned with a portion of the nucleotide sequence of *LLGL* cDNA (GenBank Accession No. D50550) (Koyama *et al.*, 1996) extending from the *Not*I site at 1446 in D50550. A short intron, which follows the GT/AG rule, is present in this region of *LLGL*. The N at nucleotide 86 in the 9-kb *Not*I sequence appears to be G in the trace file.

```
- LLGL           AAGAGATATTAAATGCTTTTATTTTCAATATTAAAAA.TCAG
                 ||||||||||||||||||||||||||||||||||||| ||||
+ FLII   TGCCCTTTAAGAGATATTAAATGCTTTTATTTTCAATATTAAAAA.TCAG  14101
         ||   ||||| ||||||||||||||||||||||||||||||||||| ||||
+ Fliih  TGTAGTTTAAAAGATATTAAATGCTTTTATTTTCAATATTAAAAAATCAG
            |||||||||||||||||||||||||||||||||||||||||||||||
- Llglh     AGTTTAAAAGATATTAAATGCTTTTATTTTCAATATTAAAAAATCAG   4218

- LLGL   TATTTTTAATATTAAAATGGCGCTAATTTT
         ||||||||||||||||||||||||||||||
+ FLII   TATTTTTAATATTAAAATGGCGCTAATTTT                      14131
         || ||||||||||||||   |     |||
+ Fliih  TACTTTTAATATTAAAACAACTAGTTTTTAAAAAAAGACAGAAAAGCCGA
         ||||||||||||||||||||||||||||||||||||||||||||||||||
- Llglh  TACTTTTAATATTAAAACAACTAGTTTTTAAAAAAAGACAGAAAAGCCGA   4168
```

FIG. 7. Region of overlap of human *FLII* and *LLGL* and mouse *Fliih* and *Llglh* genes. Poly(A) signals and sites for which polyadenylated cDNAs have been observed are indicated by shaded text and black boxes, respectively. (+) indicates that the gene is on the (+) strand and (−) indicates that the gene is on the (−) strand.

that *CAPG* arose by loss of domain 2 from a gelsolin-like dimeric precursor gene (Claudianos and Campbell, 1995), based on phylogenetic analysis of family members. Evidence presented here strongly supports this conclusion. First, analysis of the exon/intron structure of family members indicates that Cap G is more closely related to domain 1 of the dimeric proteins than domain 2. A prediction from this is that the intron positions in Cap G will be found more similar to those of domains 1 than to domains 2 of gelsolin and adseverin when their complete genomic structures are available. Second, phylogenetic analysis of the separated domains 1 and 2 of family members (Fig. 5) strongly supports the close relationship of Cap G to domain 1 of gelsolin and adseverin. The duplication event must have occurred before the latest common ancestor of higher eukaryotes and *Dictyostelium* prior to 1.1 billion years ago, since *Dictyostelium* contains protovillin (Hoffman *et al.*, 1993). We find no evidence for gelsolin-related sequences in the complete yeast genome, and the most closely related LRR sequence is that of adenylate cyclase (unpublished results). It is tempting to speculate that some of the positionally conserved introns (Fig. 4) may contain regulatory elements as proposed by Mattick (1994).

The gelsolin-like domain of family members is involved in the capping and severing of actin filaments (Hartwig and Kwiatkowski, 1991). Cap G caps actin filaments, but has no severing activity, although this can be restored by changing small divergent portions of the Cap G sequence back to the corresponding residues of gelsolin domain 1 (Southwick, 1995). It has recently been shown using expressed material that the gelsolin-like domain of the human FLII protein binds actin (Orloff *et al.*, 1995).

The presence of an actin-binding domain and a domain that may bind a ras-like molecule (Campbell *et al.*, 1993; Claudianos and Campbell, 1995), combined with the phenotype of homozygous lethal mutations in *Drosophila fliI*, suggests that the fliI protein is involved in rearrangements of the actin cytoskeleton during cellularization in early embryogenesis, possibly involving ras family member-mediated cytoskeletal regulatory processes. The fliI protein is also involved in the organization of indirect flight muscle (Deak *et al.*, 1982; Miklos and de Couet, 1990). Presumably the human FLII protein is involved in functionally analogous processes. In this respect it is interesting that the human *FLII* gene is most highly expressed in muscle.

The *FLII* gene was the first protein coding gene mapped into the critical region deleted in SMS (Chen *et al.*, 1995), a relatively common (⩾1 in 25,000 live births) microdeletion syndrome with a wide range of physical, developmental, functional, and behavioral effects (Chen *et al.*, 1995). Other genes mapping to this region include *LLGL*, the human homologue of the *D. melanogaster lethal(2) giant larvae* gene (*l(2)gl*) (Strand *et al.*, 1995; Koyama *et al.*, 1996). We have shown that the 3′ end of *LLGL* overlaps the 3′ end of *FLII* and that the 3′ end of the mouse homologue *Llglh* (Tomotsune *et al.*, 1993) overlaps the 3′ end of *Fliih*. *LLGL* and *FLII* are in the opposite transcriptional orientation in a tail-to-tail arrangement (Figs. 2 and 7). Since *Llglh* maps to mouse chromosome 11 (Kuwabara *et al.*, 1994), *Fliih* must also map there. In *D. melanogaster*, *l(2)gl* maps to 21A, whereas *fliI* maps to 19F, and in *C. elegans*, the homologue of *l(2)gl*, F56F10.4 (GenBank Accession No. U51993), maps to the X chromosome, whereas the *fliI* homologue maps to chromosome III (Campbell *et al.*, 1993; Sulston *et al.*, 1992). 3′ overlaps of mammalian genes are apparently rare (Tee *et al.*, 1995; Bristow *et al.*, 1993). The significance of the present instance of overlapping genes is unclear, although the conservation of the overlap region between human and mouse suggests the possibility of conserved function and also suggests that mutations in this region could affect expression of both genes. In this context, it is intriguing that the LLGL protein interacts with nonmuscle myosin II in a cytoskeletal network (Strand *et al.*, 1995), while the FLII protein interacts with cytoskeletal actin (Campbell *et al.*, 1993; Orloff *et al.*, 1995).

The size of the *FLII* mRNA from Northern analysis is 4.4 kb, in agreement with the composite cDNA size of 4142 bp, allowing for a poly(A) tail of ~100–200 bp. It appeared likely from alignments with *C. elegans* and *D. melanogaster fliI* sequences that two bases of the ATG translation initiation codon were missing from the *FLII* cDNA sequence (Campbell *et al.*, 1993). An EST clone, GenBank Accession No. R33910, extends 37 bp upstream and, with the genomic sequence, confirms the ATG initiation codon. A potential TATA box and a CpG island (Cross *et al.*, 1994), indicative of a 5′ regulatory region, occur just further 5′ to this point.

The complete sequence of the human *FLII* gene will be of utility in the analysis of human genetic disorders due to mutations in this conserved gene. These may include muscle disorders, based on the muscle phenotype of the viable *Drosophila* alleles and the elevated level of expression of *FLII* in human skeletal muscle, and could include some features of SMS (Chen *et al.*, 1995), if mutations in the *FLII* gene on the nondeleted chromosome 17 contribute to the phenotype. Alternatively, some SMS phenotypic features may result from *FLII* haploinsufficiency via the common deletion at 17p11.2 (Chen *et al.*, 1995). If this is the case, loss-of-function mutations in *FLII* could also generate some SMS-like phenotypic features. Mouse *Fliih* is of similar size to the human *FLII* gene (unpublished results), and gene targeting using the 15-kb mouse sequence is planned. The generation of altered alleles of *Fliih* may provide important insight into the biological role of the FLII protein and into whether haploinsufficiency at the *FLII* locus contributes to the phenotype of SMS.

ACKNOWLEDGMENTS

We thank A. Gibbs for support, C. McCrae and P. Milburn of the ANU Biomolecular Resource Facility for running the sequencers, A. Lewis, T. Goodsell, and J. Whitehead for illustrations, and J. Wilson and M. Whittaker for photography.

REFERENCES

Bazari, W. L., Matsudaira, P., Wallek, M., Smeal, T., Jakes, R., and Ahmed, Y. (1988). Villin sequence and peptide map identify six homologous domains. *Proc. Natl. Acad. Sci. USA* **85:** 4986–4990.

Bristow, J., Tee, M. K., Gitelman, S. E., Mellon, S. H., and Miller, W. L. (1993). Tenascin-X: A novel extracellular matrix protein encoded by the human XB gene overlapping P450c21B. *J. Cell Biol.* **122:** 265–278.

Campbell, H. D., Schimansky, T., Claudianos, C., Ozsarac, N., Kasprzak, A. B., Cotsell, J. N., Young, I. G., de Couet, H. G., and Miklos, G. L. G. (1993). The *Drosophila melanogaster flightless-I* gene involved in gastrulation and muscle degeneration encodes gelsolin-like and leucine-rich-repeat domains, and is conserved in *Caenorhabditis elegans* and human. *Proc. Natl. Acad. Sci. USA* **90:** 11386–11390.

Chen, K.-S., Gunaratne, P. H., Hoheisel, J. D., Young, I. G., Miklos, G. L. G., Greenberg, F., Shaffer, L. G., Campbell, H. D., and Lupski, J. R. (1995). The human homologue of the *Drosophila melanogaster flightless-I* gene (*fliI*) maps within the Smith–Magenis microdeletion critical region in 17p11.2. *Am. J. Hum. Genet.* **56:** 175–182.

Claudianos, C., and Campbell, H. D. (1995). The novel *flightless-I* gene brings together two gene families, actin binding proteins related to gelsolin and leucine-rich-repeat proteins involved in ras signal transduction. *Mol. Biol. Evol.* **12:** 405–414.

Cross, S. H., Charlton, J. C., Nan, X., and Bird, A. P. (1994). Purification of CpG islands using a methylated DNA binding column. *Nature Genet.* **6:** 236–244.

de Couet, H. G., Fong, K. S. K., Weeds, A. G., McLaughlin, P. J., and Miklos, G. L. G. (1995). Molecular and mutational analysis of a gelsolin-family member encoded by the *flightless I* gene of *Drosophila melanogaster. Genetics* **141:** 1049–1059.

Deak, I. I., Bellamy, P. R., Bienz, M., Dubuis, Y., Fenner, E., Gollin, M., Rähmi, A., Ramp, T., Reinhardt, C. A., and Cotton, B. (1982). Mutations affecting the indirect flight muscles of *Drosophila melanogaster. J. Embryol. Exp. Morphol.* **69:** 61–81.

Deininger, P. (1983). Random subcloning of sonicated DNA: Application to shotgun DNA sequence analysis. *Anal. Biochem.* **129:** 216–223.

Devereux, J., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12:** 387–395.

Felsenstein, J. (1993). PHYLIP—Phylogeny inference package, University of Washington, Seattle, WA.

Hartwig, J., and Kwiatkowski, D. (1991). Actin-binding proteins. *Curr. Opin. Cell Biol.* **3:** 87–97.

Hoffman, A., Noegel, A. A., Bomblies, L., Lottspeich, F., and Schleicher, M. (1993). The 100kDa F-actin capping protein of *Dictyostelium* amoebae is a villin prototype ('protovillin'). *FEBS Lett.* **328:** 71–76.

Homyk, T., Jr., and Sheppard, D. E. (1977). Behavioral mutants of *Drosophila melanogaster.* I. Isolation and mapping of mutations which decrease flight ability. *Genetics* **87:** 95–104.

Kallioniemi, O.-P., Kallioniemi, A., Mascio, L., Sudar, D., Pinkel, D., Deaven, L., and Gray, J. (1994). Physical mapping of chromosome 17 cosmids by fluorescence *in situ* hybridization and digital image analysis. *Genomics* **20:** 125–128.

Kobe, B., and Deisenhofer, J. (1995). Proteins with leucine-rich repeats. *Curr. Opin. Struct. Biol.* **5:** 409–416.

Koyama, K., Fukushima, Y., Inazawa, J., Tomotsune, D., Takahashi, N., and Nakamura, Y. (1996). The human homologue of the murine *Llglh* gene (*LLGL*) maps within the Smith–Magenis syndrome region in 17p11.2. *Cytogenet. Cell Genet.* **72:** 78–82.

Kuwabara, K., Takahashi, Y., Tomotsune, D., Takahashi, N., and Kominami, R. (1994). mgl-1, a mouse homologue of the *Drosophila* tumor-suppressor gene *l(2)gl*, maps to chromosome 11. *Genomics* **20:** 337–338.

Kwiatkowski, D. J., Mehl, R., and Yin, H. L. (1988). Genomic organization and biosynthesis of secreted and cytoplasmic forms of gelsolin. *J. Cell Biol.* **106:** 375–384.

Mahajan-Miklos, S., and Cooley, L. (1994). The villin-like protein encoded by the *Drosophila quail* gene is required for actin bundle assembly during oogenesis. *Cell* **78:** 291–301.

Mattick, J. (1994). Introns: Evolution and function. *Curr. Opin. Cell Biol.* **4:** 823–831.

Miklos, G. L. G., and de Couet, H. G. (1990). The mutations previously designated as *flightless-I*^{3}, *flightless-O*^{2} and *standby* are members of the *W-2* lethal complementation group at the base of the X-chromosome of *Drosophila melanogaster. J. Neurogenet.* **6:** 133–151.

Mishra, V. S., Henske, E. P., Kwiatkowski, D. J., and Southwick, F. S. (1994). The human actin-regulatory protein Cap G: Gene structure and chromosome location. *Genomics* **23:** 560–565.

Orloff, G. J., Allen, P. G., Miklos, G. L. G., Young, I. G., Campbell, H. D., and Kwiatkowski, D. J. (1995) Human flightless-I has actin binding ability. *Mol. Biol. Cell* **6:** 139.

Perrimon, N., Smouse, D., and Miklos, G. L. G. (1989). Developmental genetics of loci at the base of the *X* chromosome of *Drosophila melanogaster. Genetics* **121:** 313–331.

Pringault, E., Robine, S., and Louvard, D. (1991). Structure of the human villin gene. *Proc. Natl. Acad. Sci. USA* **88:** 10811–10815.

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

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). "Molecular Cloning: A Laboratory Manual," 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Southwick, F. S. (1995). Gain-of-function mutations conferring actin-severing activity to human macrophage Cap G. *J. Biol. Chem.* **270:** 45–48.

Staden, R. (1987). Computer handling of DNA sequencing projects. *In* "Nucleic Acid and Protein Sequence Analysis: a Practical Approach" (M. J. Bishop and C. J. Rawlings, Eds.), pp. 173–217, IRL Press, Oxford.

Strand, D., Unger, S., Corvi, R., Hartenstein, K., Schenkel, H., Kalmes, A., Merdes, G., Neumann, B., Krieg-Schneider, F., Coy, J. F., Poustka, A., Schwab, M., and Mechler, B. M. (1995). A human homologue of the *Drosophila* tumour suppressor gene *l(2)gl* maps to 17p11.2–12 and codes for a cytoskeletal protein that associates with nonmuscle myosin II heavy chain. *Oncogene* **11:** 291–301.

Sulston, J., Du, Z., Thomas, K., Wilson, R., Hillier, L., Staden, R., Halloran, N., Green, P., Thierry-Mieg, J., Qiu, L., Dear, S., Coulson, A., Craxton, M., Durbin, R., Berks, M., Metzstein, M., Hawkins, T., Ainscough, R., and Waterston, R. (1992). The *C. elegans* genome sequencing project: A beginning. *Nature* **356:** 37–41.

Swofford, D. (1993). PAUP: Phylogenetic analysis using parsimony, Version 3.1, Illinois Natural History Survey, Champaign, IL.

Tee, M. K., Thomson, A. A., Bristow, J., and Miller, W. L. (1995). Sequences promoting the transcription of the human XA gene overlapping P450c21A correctly predict the presence of a novel, adrenal-specific, truncated form of tenascin-X. *Genomics* **28:** 171–178.

Tomotsune, D., Shoji, H., Wakamatsu, Y., Kondoh, H., and Takahashi, N. (1993). A mouse homologue of the *Drosophila* tumour-suppressor gene *l(2)gl* controlled by Hox-C8 in vivo. *Nature* **365:** 69–72.

Way, M., and Weeds, A. (1988). Nucleotide sequence of pig plasma gelsolin. Comparison of protein sequence with human gelsolin and other actin-severing proteins shows strong homologies and evidence for large internal repeats. *J. Mol. Biol.* **203:** 1127–1133.

Witke, W., Sharpe, A. H., Hartwig, J. H., Azuma, T., Stossel, T. P., and Kwiatkowski, D. J. (1995). Hemostatic, inflammatory, and fibroblast responses are blunted in mice lacking gelsolin. *Cell* **81:** 41–51.

Zehetner, G., and Lehrach, H. (1994). The Reference Library System—sharing biological material and experimental data. *Nature* **367:** 489–491.