J Mol Evol (2001) 53:703–710
DOI: 10.1007/s002390010257

JOURNAL OF MOLECULAR EVOLUTION

© Springer-Verlag New York Inc. 2001

Intron Dynamics and the Evolution of Integrin β-Subunit Genes: Maintenance of an Ancestral Gene Structure in the Coral, *Acropora millepora*

Dasan M. Schmitt, Danny L. Brower

Department of Molecular and Cellular Biology, Life Sciences South Building, 1007 East Lowell Street, University of Arizona, Tucson, AZ 85721, USA

Received: 5 March 2001 / Accepted: 17 May 2001

Abstract. We have determined the genomic structure of an integrin β-subunit gene from the coral, *Acropora millepora.* The coding region of the gene contains 26 introns, spaced relatively uniformly, and this is significantly more than have been found in any integrin β-subunit genes from higher animals. Twenty-five of the 26 coral introns are also found in a β-subunit gene from at least one other phylum, indicating that the coral introns are ancestral. While there are some suggestions of intron gain or sliding, the predominant theme seen in the homologues from higher animals is extensive intron loss. The coral baseline allows one to infer that a number of introns found in only one phylum of higher animals result from frequent intron loss, as opposed to the seemingly more parsimonious alternative of isolated intron gain. The patterns of intron loss confirm results from protein sequences that most of the vertebrate genes, with the exception of β4, belong to one of two β subunit families. The similarity of the patterns within each of the β1,2,7 and β3,5,6,8 groups indicates that these gene structures have been very stable since early vertebrate evolution. Intron loss has been more extensive in the invertebrate genes, and obvious patterns have yet to emerge in this more limited data set.

Key words: Intron evolution — Integrin evolution — Vertebrate genome.

Correspondence to: Danny L. Brower; *email:* dbrower@u.arizona.edu

Introduction

Integrins are a large family of cell surface receptors and serve as the primary molecules by which metazoan cells interact with extracellular matrix (Hynes 1992). A functional integrin is composed of an α–β heterodimer, and both α and β subunits are large (approximately 100 kDa or more) transmembrane proteins. Integrin genes have been identified in all types of metazoans, including sponges and cnidarians (Pancer et al. 1997; Brower et al. 1997; Wimmer et al. 1999), but they have not been found in any fungi or plants, including the completed genomes of *Saccharomyces cerevisiae* and *Arabidopsis thaliana* (Mewes et al. 1997; Arabidopsis Genome Initiative 2000). Integrin α and β subunits are nonhomologous to one another, but for each subunit type the overall structural properties of the proteins appear to be highly conserved. Indicative of this structural conservation, most β subunits contain 56 extracellular cysteines, proposed to form a set of conserved disulfide bonds. One likely reason for the overall high degree of conserved structure is that the α and β subunits appear to interact along their entire lengths and undergo concerted allosteric changes when the heterodimer switches between active and inactive conformations (Humphries 1996).

We set out to examine the genomic organization of an integrin β-subunit gene, corresponding to cDNAs previously isolated from the coral, *Acropora millepora* (Brower et al., 1997). We found an unexpectedly large number of introns in the cnidarian gene and compared the locations of these to β-subunit genes from *Caenorhabditis elegans, Drosophila melanogaster,* and humans. Integrin β subunits seem to be well suited to this

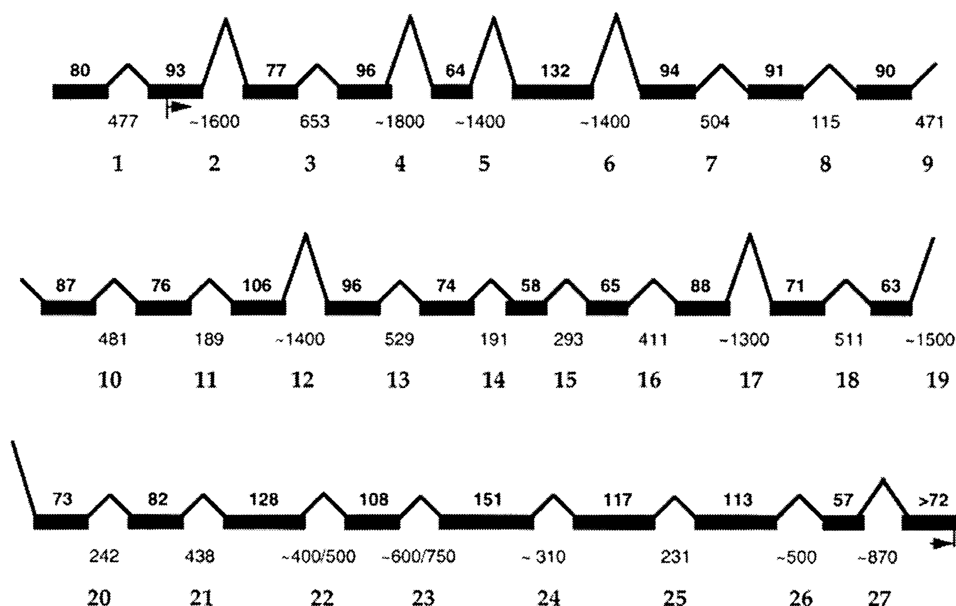

Fig. 1. Genome organization of the *A. millepora* βCn1 integrin gene. Introns are numbered 1–27, and the sizes of the introns and exons are indicated. *Arrows* indicate the positions of the initiating AUG and the stop codon. Introns 22 and 23 were polymorphic in size in our sample.

type of analysis, since they are large and structurally well conserved, containing residues throughout that permit one to make sequence homology assignments with confidence. The results lead to the conclusion that the coral introns represent an ancestral genomic structure and provide novel insights into the evolution of genomic organization in animals.

Methods

Acropora millepora genomic DNA was a kind gift of Julian Catmull and David Miller and was extracted from embryos collected near Magnetic Island, Queensland, Australia. Using the *A. millepora* integrin β-subunit cDNA sequence [Genbank accession number AF005356 (Brower et al. 1997)], PCR primers were generated from multiple sites and fragments were amplified using Pfu Turbo DNA polymerase (Stratagene) from the genomic template. Amplified fragments were sequenced directly at the LMSE DNA sequencing facility of the University of Arizona, and the splice sites of all introns within the cDNA sequence were identified. Locations in the cDNA sequence of the last nucleotide of each exon (where 1 is the start of the initiating AUG) are as follows: exon 1, nucleotide – 30; 2, 64, 3, 41; 4, 237; 5, 301; 6, 433; 7, 521; 8, 612; 9, 702; 10, 789; 11, 867; 12, 963; 13, 1059; 14, 1133; 15, 1191; 16, 1256; 17, 1344; 18, 1415; 19, 1478; 20, 1552; 21, 1633; 22, 1761; 23, 1869; 24, 2020; 25, 2137; 26, 2250; and 27, 2307.

Sequences of other genes for comparisons were from work in our lab and the *C. elegans* and *Drosophila* genomic sequencing projects for the invertebrate genes (*C. elegans* Sequencing Consortium 1998; Adams et al. 2000). Human gene structures were from Weitzman et al. (1991) (β2), Lanza et al. (1990) (β3), Iacovacci et al. (1997) (β4), and Jiang et al. (1992) (β7), from sequences deposited in Genbank by the International Human Genome Sequencing Consortium (2001) (primarily β1, β5, and β8, although parts of the previously published sequences were checked as well), and from the Celera.com web site for β6 (Venter et al. 2001). To identify splice sites, Blast searches were run using both protein and cDNA sequences against the appropriate genome sequences. (Genbank accession numbers for cDNA sequences: β1, NM 002211; β2, NM 000211; β3, NM 000212; β4, X52186; β5, J05633; β6, NM 000888; β7, NM 000889; β8, NM 002214; βPS, J03251; βneu, L13305; and βpat-3, U19744.) Multiple sequence alignments were done using ClustalW (Baylor College of Medicine web site), with subsequent adjustments by hand, keying on the homologous cysteines.

Results

Using sequence from an *A. millepora* cDNA as a guide, we generated primers from which we PCR-amplified sections of the corresponding gene from *A. millepora* genomic DNA. We then directly sequenced the amplified DNA to determine the locations of all introns. In all, 27 introns were identified within the boundaries of the cDNA sequence, one 5′ to the initiating methionine and 26 within the coding sequence (Fig. 1).

The *A. millepora* β-subunit gene is broken up into relatively small exons, ranging in size from 57 to 151 nucleotides. The intron sizes range from 115 base pairs (bp) to almost 2 kilobases (kb). (The sizes of some introns were not determined by complete sequencing but are estimated from gels.) Not surprisingly, intron sequences display more polymorphism than the exon sequences, and two introns from our sample show significant polymorphism in size. The sequences surrounding the splice sites match well with the consensus frequencies seen for vertebrate mRNAs (Padgett et al. 1986).

Genomic structures have been determined for the single *C. elegans,* two *Drosophila,* and eight human integrin β genes (see Methods). By way of background, the nematode and fly each contain only one typical integrin β-subunit gene (Hynes and Zhao 2000). The *Drosophila*

βν is a highly divergent protein with a restricted expression that suggests a specialized function (Yee and Hynes 1993), and there are no obvious orthologues to βν known from other animals.

Figure 2 shows the assembled intron locations, and comparisons of the different genes are summarized in Table 1. Although many of the genes, including the coral, have an intron in the 5′ untranslated sequence, we did not include this since there is no corresponding protein sequence from which to make independent inferences of positional homology. Even though absolute sequences are not always well conserved in parts of integrin β subunits, the positions of the 56 cysteines present in most of the proteins allow structural homologies to be assigned with good resolution for most of the sequence.

Of the 26 coding region coral introns, 24 have at least one cognate in a higher animal. Additionally, we find that the position of one of the two "unique" coral introns, number 10, is precisely conserved in the sponge, *Ophlitaspongia tenuis* (S. Miller and D.B., unpublished). This suggests that the coral β-subunit gene structure is ancestral (see Discussion), and we describe the results based on this interpretation. We also assume that intron positions can move slightly during evolution. Although there is some disagreement regarding the scope of such intron sliding (e.g., De Souza et al. 1998; Rzhetsky et al. 1997; Stoltzfus et al. 1997; Roy et al. 1999), for most of the analyses this seems much more parsimonious than positing multiple intron losses and gains to explain relatively minor shifts in position. Many of the slight deviations in position, especially those in the highly divergent human β4 and fly βν genes, are likely to result from uncertainties in the multiple sequence alignment. Introns that may represent genuine additions to the ancestral coral positions are the human sets near coral introns 6, 7, 18, and 23. (For easy reference, all noncoral intron positions are indicated relative to the coral positions, numbered in Fig. 2.) Since each of these coral introns is matched with a cognate in the human β4 gene, we infer that the coral introns are ancestral.

For the vertebrate introns, the distributions clearly sort into three sets, comprised of the divergent β4 and the related β1,2,7 and β3,5,6,8 groups. For the latter two groups, most intron sites are conserved for all seven genes, except for introns 6, 7, 18, and 24, where the patterns segregate the vertebrate genes into the two groups.

Two of the vertebrate genes, β4 and β8, have cytoplasmic domains that are generally considered to be nonhomologous to the other β subunits. However, each gene has a cytoplasmic intron that aligns with other β-gene introns, as determined by the distance from the transmembrane domain. The first cytoplasmic intron of β4 aligns precisely with a coral intron, even though all amino acid sequence homology disappears following the conserved tryptophan–lysine doublet at the membrane interface. Additionally, the cytoplasmic intron of β8 aligns reasonably well with the other vertebrate cytoplasmic introns.

Compared to the vertebrates the fly and worm genes have many fewer introns. All of these invertebrate introns line up well with coral cognates. There are no clear cladistic relationships between these invertebrate intron positions. In particular, the *Drosophila* βν is not obviously more similar to the fly βPS than to the nematode βpat-3 or to any of the vertebrate genes.

Discussion

The Coral Genomic Structure Is Ancestral

The relatively large number of introns found in the *Acropora* integrin β-subunit gene raised the possibility that intron accumulation may be a derived state of this gene. However, comparisons with other integrin β-subunit genes reveals that the coral pattern is ancestral and that loss in the more advanced phyla has been extensive and variable. (It should be noted that the term "ancestral" here refers to the metazoan lineage only; since integrins appear to be specific to animals, our data do not address issues related to more distant events.)

Only one of the coral introns has not been found in any of the other animals examined. Since at least nine of the coral introns (and possibly more, depending on how one interprets some potential cases of intron sliding), have been retained in only one of the other β genes, it seems reasonable statistically that loss of all of the noncoral cognates could be expected in some cases. Also, the coral gene is characterized by fairly short exons, and if the unique coral intron was deleted the resulting exon would be the longest in the sequence. These considerations suggest that the unique coral intron represents a site at which the cognates have been lost from all of the other animals, as opposed to a site at which the coral lineage gained an intron after its divergence from other metazoans.

Less clear is whether any ancestral β-subunit introns have been lost from *A. millepora.* Although some exons could be bisected to yield exon pairs that would be within the minimum size range present in the rest of the gene (50–60 bp), there are no corresponding introns in the other phyla to indicate the presence of primal introns at these sites. The best case for an intron unique to noncoral species, which might therefore indicate a site of coral loss, is the β1,2,7 intron between coral intron 6 and coral intron 7. However, it should be noted that if the exon bounded by coral intron 6 and coral intron 7 was divided in the middle at one time in an *Acropora* ancestor, the resulting exons would be the shortest in the gene.

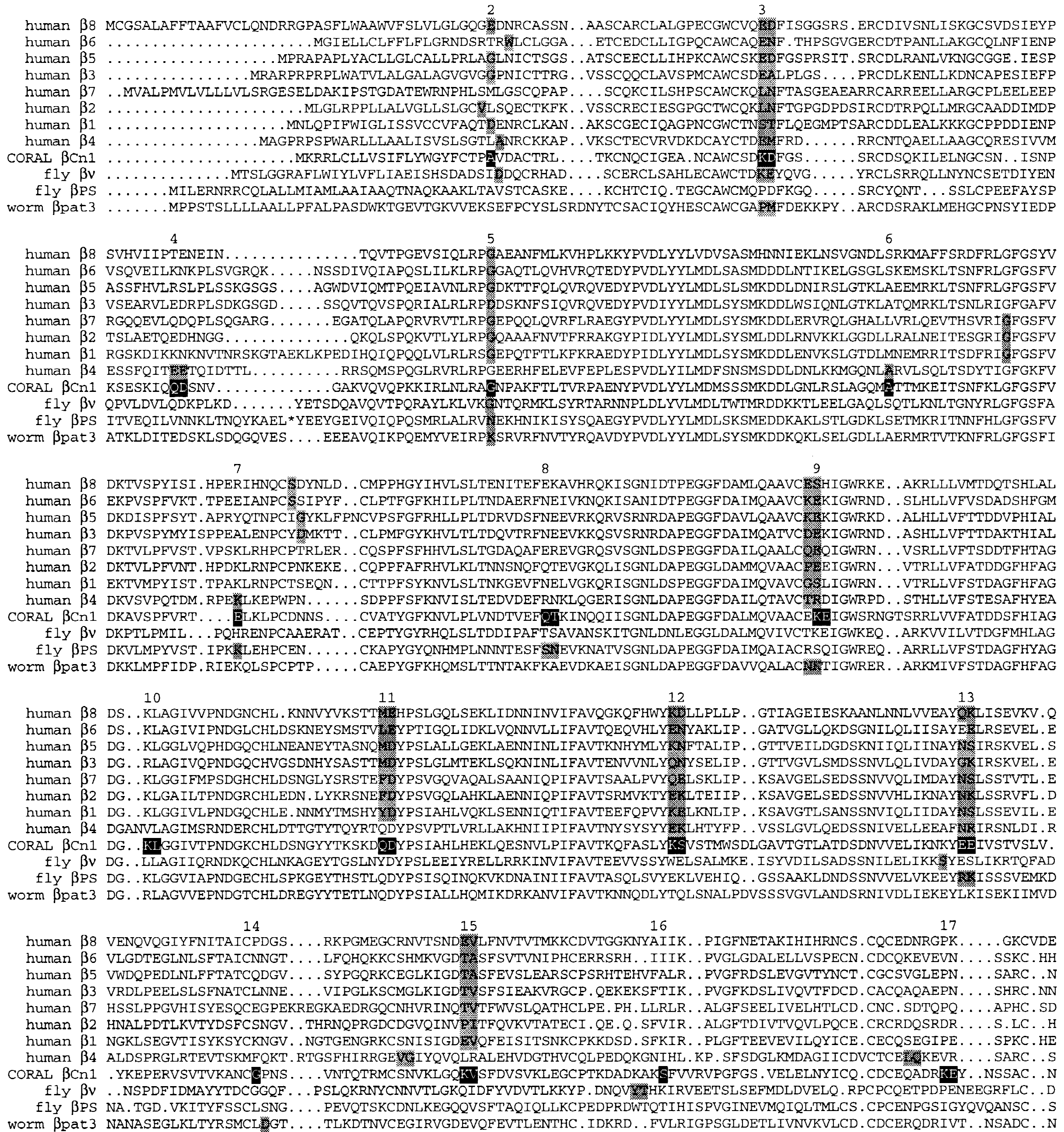

Fig. 2. Positions of introns in integrin β-subunit genes. Intron positions are indicated by *shading;* coral introns are in the *darker boxes* between the human and the invertebrate sequences and are numbered (intron 1 is in the untranslated 5′ leader). Boxes with two residues denote splice sites that fall at the codon boundary. For nucleotide positions of the coral splice sites, see Methods.

We cannot say why the coral integrin β gene should retain the ancestral introns so effectively. Previously determined cnidarian gene structures, including those of noncoral cnidarians, do not suggest any trend toward intron richness relative to animals such as vertebrates (Bosch et al. 1989; Fisher and Bode 1989; Aerne et al. 1993; Lopez et al. 1994; Gendeh et al. 1997; Sun et al. 1997; Vibede et al. 1998; Erber et al. 1999; Spafford et al. 1999; Tom et al. 1999; Miller et al. 2000). Indeed, several of these studies indicate that cnidarians can have reduced numbers of introns relative to higher animals. For example, Erber et al. (1999) examined gene structures of another structurally conserved protein family, the nuclear lamins. In this work, as with our study, most intron positions were conserved between the vertebrate genes, and the nematode and fly had reduced numbers of introns in divergent patterns. However, the cnidarian (*Hydra attenuata*) was missing the majority of the introns seen in the higher animals, retaining just three at the 3′ end of the gene.

One of the more interesting studies in this regard is that of Spafford et al. (1999), who compared intron positions in a sodium channel gene from a hydrozoan jellyfish with those from vertebrates and flies. Like the

```
                            18                           19                              20                              21
  human β8 TFLDSKCFQC......DENKCHFDEDQFSSES.......CKSHK.....DQPVCSGRGVCVCGKCSCHKIKLGK..VYGKYCEKDDFSCPYHHGNLCAG..HGEC
  human β6 GNGSFQCGVCACHPGHMGPRCECGEDMLSTDS.......C.KEA....PDHPSCSGRGDCYCGQCICHLSPYGN..IYGPYCQCDNFSCVRHKGLLCGG..NGDC
  human β5 GSGTYVCGLCECSPGYLGTRCECQDGENQSVYQNL....C.REA....EGKPLCSGRGDCSCNQCSCFESEFGK..IYGPFCECDNFSCARNKGVLCSG..HGEC
  human β3 GNGTFECGVCRCGPGWLGSQCECSEEDYRPSQQDE....C.SPR.....GQPVCSQRGECLCGQCVCHSSDFGK..ITGKYCECDDFSCVRYKGEMCSG..HGQC
  human β7 GQGHLQCGVCSCAPGRLGRLCECSVAELSSPDLESG...CRAPN....GTGPLCSGKGHCQCGRCSCSGQS......SGHLCECDDASCERHEGILCGG..FGRC
  human β2 GKGFLECGICRCDTGYIGKNCECQTQGRSSQELEGS...CRKDN.....NSIICSGLGDCVCGQCLCHTSDVPGKLIYGQYCECDTINCERYNGQVCGGPGRGLC
  human β1 GNGTFECGACRCNEGRVGRHCECSTDEVNSEDMDAY...CRKEN.....SSEICSNNGECVCGQCVCRKRDNTNEIYSGKFCECDNFNCDRSNGLICGG..NGVC
  human β4 FNGDFVCGQCVCSEGWSGQTCNCSTGSLSDIQP......CLREG.....EDKPCSGRGECQCGHCVCYGEGRY....EGQFCEYDNFQCPRTSGFLCND..RGRC
CORAL βCn1 KLGALTCGLCACNEGRFGKFCQCDTPFSKTEQDK.....CKSSN...STDEPLCSGRGECACGECVCRVEQGQR..FYGKLCECNDFSCPEYEGNLCGCAERGVC
    fly βv YKGYLYCGMCECDEGWTGTYCNCPTDATNVTSNEALLQKCRQPFSDKSTSELVCSNHGDCDCGTCLCDPGYT......GPFCECRE..C.....LDCDE.KLADC
   fly βPS GHGTSMCGICNCDDSYFGNKCECSATDLTSKFANDTS..CRADS....TSTTDCSGRGHCVCGACECHKRPNPIEIISGKHCECDNFSCERNRNQLCSGPDHGTC
worm βpat3 G.GDMVCGVCRCKGGNVGKYCECNRPGMSTAALNEK...CKRTN.....ESAICEGRGVCNCGRCECNPRANPEEQISGEFCECDNFNCPRHDRKICAE..HGEC

                                                    22                                             23
  human β8 EAGRCQCFSGWEGDRCQ...CPSAA.AQHCVNS.KGQVCSG..RGTCV.....CGRCECTDP....RSIGRFCEHC.PTCY.TACKENWNCMQCLHPHNLSQAIL
  human β6 DCGECVCRSGWTGEYCN...CTTST..DSCVSE.DGVLCSG..RGDCV.....CGKCVCTNP....GASGPTCERC.PTCG.DPCNSKRSCIECHLSAAG.QA.G
  human β5 HCGECKCHAGYIGDNCN...CSTDI..STCRGR.DGQICSE..RGHCL.....CGQCQCTEP....GAFGEMCEKC.PTCP.DACSTKRDCVECLLLHSGKPD.N
  human β3 SCGDCLCDSDWTGYYCN...CTTRT..DTCMSS.NGLLCSG..RGKCE.....CGSCVCIQP....GSYGDTCEKC.PTCP.DACTFKKECVECKKFDREPYMTE
  human β7 QCGVCHCHANRTGRACE...CSGDM..DSCISP.EGGLCSG..HGRCK.....CNRCQCLD.....GYYGALCDQC.PGCK.TPCERHRDCAECGAFRTGPLA..
  human β2 FCGKCRCHPGFEGSACQ...CERTT..EGCLNP.RRVECSG..RGRCR.....CNVCEC.HS....GYQLPLCQEC.PGCP.SPCGKYISCAECLKFEKGPFG..
  human β1 KCRVCECNPNYTGSACD...CSLDT..STCEAS.NGQICNG..RGICE.....CGVCKCTDP....KFQGQTCEMC.QTCL.GVCAEHKECVQCRAFNKGEKK..
  human β4 SMGQCVCEPGWTGPSCD...CPLSN..ATCIDS.NGGICNG..RGHCE.....CGRCHCHQQS...LYTDTICEINYSAIHPGLCEDLRSCVQCQAWGTGEKK.G
CORAL βCn1 RCRKCQCKDKYHGDACDQKNCTFFPPETICKQDAKSEMCGGADRGRCVKDSVNCYKCQCNKE.....FDGTYCENC.PNCENGMCTRNVDCALCSTFQGKSLK..
    fly βv FCGQCVCKYGWSGSKCN...CDGDT..DACVGP.TGEICSE..RGTCQ.....CEECQCEEP.....YLGKFCEID.PEKDNKLCLFYEPCVTCLIEQKQGMG..
   fly βPS ECGRCKCKPGWTGSNCG...CQESN..DTCMPPGGGEICSG..HGTCE.....CGVCKCTVNDQG.RFSGRHCEKC.PTCS.GRCQELKDCVQCQMYKTGELKNG
worm βpat3 NCGKCICAPGWTGRACE...CPIST..DSCLSA.NGKICNG..KGECI.....CGRCRCFDSPDGNRYSGAKCEIC.PTCP.TKCVEYKNCVMCQQWQTGPL.NE

                                            24                                              25
  human β8 DQCKTS...CA.............................................LMEQQHYVDQTSKCFSSPSYLRIFFIIFIVTFLIGLLKVLIIRQV
  human β6 EECVDK...CKLAGA....  TISEEEDFSKDGSVSCSLQGEN....ECLITFLITT.DNEGKTIIHSINE.KDCPKPPNIPMIMLGVSLATLLIGVVLLCIWKLL
  human β5 QTCHSL...CRDEVI.....TWVDTIVKDDQEAVLCFYKTAK....DCVMMFTYVE.LPSGKSNLTVLRE.PECGNTPNAMTILLAVVGSILLVGLALLAIWKLL
  human β3 NTCNRY...CRDEIE......SVKELKDTGKDAVNCTYKNED....DCVVRFQYYE.DSSGKSILYVVEE.PECPKGPDILVVLLSVMGAILLIGLAALLIWKLL
  human β7 TNCSTA...CAHTNV.........TLALAPILDDGWCKERTLDN...Q.LFFFLVEDDARGTVVL.RVRPQ.EKGADHTQ..AIVLGCVGGIVAVGLGLVLAYRLS
  human β2 KNCSAA...CPGLQL...........SNNPVKGRTCKERDSE....GCWVAYTLEQQDGMDRYLIYVDES.RECVAGPNIAAIVGGTVAGIVLIGILLLVIWKAL
  human β1 DTCTQE...CSYFNITKVESRDKLPQPVQPDPVSHCKE.KDVD...DCWFYFTYSVNGNNEVM.VHVVEN.PECPTGPDIIPIVAGVVAGIVLIGLALLLIWKLL
  human β4 RTCEE....CNFKVK......MVDELKRAEEVVVRCSFRDEDD...DCTYSYTMEGDGAPGPNSTVLVHKKKDCP.PGSFWWLIPLLLLLLPLLALLLLLCWKYC
CORAL βCn1 .ECKQSGK.CEENVL..EVQIVDDIKKKTDEGLYRCEGIDGQD...GCTYYFTTETEADSKNFVLYVQKDKVSCPTEAPVLPIVLGVVGGILFLGLLILIVIKGL
    fly βv .VCENLTEICSSLDR..QETYPYNFVHELDPGQDQCLVRLVNKHGIQCDSFFAYQVI.DHSNFLTIQA...VDC.EPPDYVALVGYISAFTLLIGLLIIFIILWY
   fly βPS DDCARN...CTQFVP....VGVEKVEIDETKDEQMCKF.FDED...DCKFMFKYSEQGELHVYAQEN....KECPAKVFMLGIVMGVIAAIVLVGLAILLLWKLL
worm βpat3 TACDQ....CEFKVI.......PVEELPNLNETTPCQFVDPAD...DCTFYYLYYYDEATDNATVWVRKH.KDCPPPVPVLAIVLGVIAGIVILGILLLLLLWKLL

                26                   27
  human β8 ILQWNSNKIKSSSDYRVSASKKDKLILQSVCTRAVTYRREKPEEIKMDISKLNAHETFRCNF
  human β6 VSFHDRKEVAKFEAERSKAKWQTGTNPLYRGSTSTFKNVTYKHREKQKVDLSTDC
  human β5 VTIHDRREFAKFQSERSRARYEMASNPLYRKPISTHTVDFTFNKFNKSYNGTVD
  human β3 ITIHDRKEFAKFEEERARAKWDTANNPLYKEATSTFTNITYRGT
  human β7 VEIYDRREYSRFEKEQQQLNWKQDSNPLYKSAITTTINPRFQEADSPTL
  human β2 IHLSDLREYRRFEKEKLKSQWNND.NPLFKSATTTVMNPKFAES
  human β1 MIIHDRREFAKFEKEKMNAKWDTGENPIYKSAVTTVVNPKYEGK
  human β4 ACCKACLALLPCCNRGHMVGFKEDHYMLRENLMASDHLDTPMLRSGNLKGRDVVRWKVTNNMQRPGF + hundreds more
CORAL βCn1 FTMVDRIEYQKFERERMHSKWTREKNPLYQAAKTTFENPTYAGGRQ
    fly βv IRAKDAREYAKFEEDQKNS..VRQENPIYRDPVGRYEVPKALSVKYDENPFAS
   fly βPS TTIHDRREFARFEKERMNAKWDTGENPIYKQATSTFKNPMYAGK
worm βpat3 TVLHDRSEYATFNNERLMAKWDTNENPIYKQATTTFKNPVYAGKAN
```

Fig. 2. Continued.

Table 1. Intron positions shared by integrin β-subunit genes[a]

	Total	βCn1	β4	β1,2,7	β3,5,8	βPS	βν
Coral βCn1	26						
Human β4	17	17					
Human β1,2,7	14	11	8				
Human β3,5,8	14	12	9	12			
Fly βPS	5	5	2	3	3		
Fly βν	5	5	3	4	4	2	
Worm βpat-3	7	7	3	4	5	2	2

[a] The table includes only introns in the coding region. It does not include possible vertebrate cognates near coral introns 6, 7, 18, and 23. It does include the β4 intron near coral intron 15, as this apparent misalignment is likely to result from problems with the multiple sequence alignment, due to β4 nonhomology in this region. No β4 introns after position 26 are included, since the extensive cytoplasmic domain of this gene is nonhomologous to that of other integrins.

integrin β subunit, these genes have a highly conserved overall structure, in this case consisting of four repeats of six transmembrane domains. They found that 17 intron positions were conserved between the jellyfish and two vertebrate genes, including all 11 present in regions encoding the conserved domains within the membrane. Thus, although the cnidarian was not conspicuously intron rich, there appears to have been very little loss or gain in any of the animals examined for the most conserved structural domains of the proteins.

Anthozoans (such as corals) occupy a basal position in the phylum Cnidaria (Schuchert 1993; Bridge et al. 1995), and it will be interesting to see if corals generally retain ancestral genomic structures more faithfully than other cnidarians or if the integrin gene is unusual in this respect. From the relatively few *Acropora* or other coral genes that have been characterized previously for genomic structure, there has been no indication of unusual intron richness (Tom et al. 1999; Miller et al. 2000).

Metazoan Intron Loss Has Been Extensive

There is an ongoing debate as to whether the evolution of gene structures is characterized primarily by intron loss

or intron gain, with the complication that introns can occasionally move or "slide" to some degree (DeSouza et al. 1998; Rzhetsky et al. 1997; Stoltzfus et al. 1997; Roy et al. 1999; Gilbert 1987; Palmer and Logsdon 1991; Stoltzfus et al. 1994). Tests of these ideas are limited by the fact that all genes, including those of primitive organisms, have been evolving for many millions of years since the divergence of the major phylogenetic groups, and the ancestral gene structure can be difficult to infer. Because of this, investigators often are forced to rely on statistical analyses of relatively small data sets, and it is not unheard of for different studies to reach dissimilar conclusions from similar data (e.g., Dibb and Newman 1989; Nyberg and Cronhjort 1992).

The remarkable correspondence of almost all of the introns in the intron-rich coral integrin β-subunit gene with one or more introns in a homologue from a higher animal allows one to view the coral gene as an ancestral "baseline." Although suggestions of intron gain and sliding are evident in Fig. 2, it is clear that evolution of the gene structure in the higher animal homologues is characterized predominantly by the loss of ancestral introns. It is particularly noteworthy that without the coral baseline sequence, at least 15 of the *C. elegans, Drosophila,* or human introns would be restricted to a single phylum, and based on simple parsimony arguments these would likely be considered as evidence for sporadic intron gain, as opposed to extensive intron loss. In metazoans at least, intron loss may have been much more extensive than previously suspected, and this must be taken into account when making estimations of parsimony.

The Vertebrate Introns Have Been Stable Since the Early Gene Duplications

An early origin for the gene duplications that created the β1,2,7 and β3,5,6,8 groups is indicated by the fact that members of both groups have been identified in an amphibian (Ransom et al. 1993). Within the human β1,2,7 and β3,5,6,8 groups, there are 26 locations in which all three or four members of a group are found, and only two locations in which one member of a group is absent. In one of these two sites (intron 24), there is an extensive gap in the protein sequence for β8, and the intron has disappeared along with the coding sequence. So, discounting this deletion, there is only one case of a single intron loss within a group, intron 2 of β7. Even this intron is in the N-terminal unconserved domain (including the signal sequence), and so it is possible that the intron was not lost precisely, but disappeared along with a replacement of the coding sequence due to a larger rearrangement affecting the gene.

The above observations indicate that most of the intron loss in the human β-subunit genes occurred relatively early in the evolution of the phylum and that these genes have been very stable with respect to intron loss

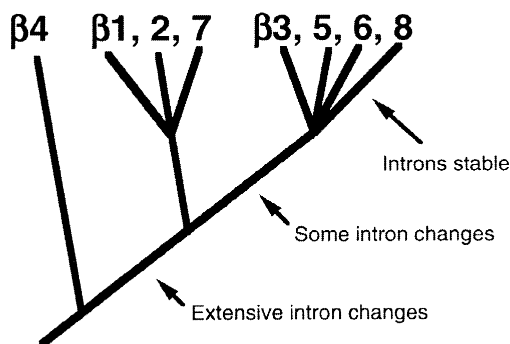

Fig. 3. Groups of vertebrate integrin β subunits. The extent of intron remodeling decreases with time and is essentially zero in the most recent branches, which are hundreds of millions of years old.

for hundreds of millions of years. It will be very interesting to see if this is a general property of human and other vertebrate genomes. A similar conclusion cannot be drawn for the invertebrate genes; to address this question, one would like to have genomic structures for a large sample of integrin β genes from one phylum.

Integrin β-Gene Evolution

The β1,2,7 and β3,5,6,8 genes group together based on protein sequence (Brower et al. 1997; Hughes 1992; Burke 1999; Hughes 2001), as well as intron pattern. However, the β4 protein sequence does not fit neatly into either the β1,2,7 or the β3,5,6,8 group. The β4 sequence is also difficult to interpret from an evolutionary standpoint since this protein is functionally unique, being the only vertebrate β subunit that contributes to an integrin that associates with intermediate filaments (Van der Neut et al. 1996; Rezniczek et al. 1998). Thus, the divergent protein sequence could reflect an unusual rate or direction of selection.

The β4 intron pattern indicates conclusively that this gene diverged early, before the separation of the other vertebrate groups. (Fig. 3) Indeed, the pattern and sequence of β4 are so unusual that it raises the question of whether this gene diverged before the establishment of a precursor vertebrate lineage (see also Hughes 2001). There is no β4 orthologue in *Drosophila* or *C. elegans,* but we cannot rule out the possibility that an ancestral β4 gene was lost from these lineages. This is especially true since, although *C. elegans* does express intermediate filament genes (Dodemont et al. 1994), the *Drosophila* genome appears to be missing such genes completely (Goldstein and Gunawardena 2000).

The invertebrate β genes display no clear pattern of intron loss that would shed light on their cladistic relationships. In particular, the fly βPS and βν do not show any similarities that might allow one to determine if the βν gene, which so far is unique to flies, was initially established in the arthropod lineage. This may suggest a deeper derivation for βν, but it also is possible that the

patterns are less clear because of a generally high rate of intron loss in the ecdysozoa.

Acknowledgments. We thank Julian Catmull and David Miller for the genomic DNA, Tom Bunch and Alison Jannuzi for guidance, and Brian and his gang for making sequencing painless. David Miller and Rick Hallick provided helpful comments on the manuscript, and Rachel Wenzel helped put it together. This work was supported by a grant from the NIH and the Undergraduate Biology Research Program of the University of Arizona.

References

Adams MD, et al. (2000) The genome sequence of *Drosophila melanogaster.* Science 287:2185–2195

Aerne BL, Schmid V, Schuchert P (1993) Actin-encoding genes of the hydrozoan *Podocoryne carnea.* Gene 131:183–192

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana.* Nature 408: 796–815

Bosch TC, Unger TF, Fisher DA, Steele RE (1989) Structure and expression of STK, a src-related gene in the simple metazoan Hydra attenuata. Mol Cell Biol 9:4141–4151

Bridge D, Cunningham CW, DeSalle R, Buss LW (1995) Class-level relationships in the phylum cnidaria: Molecular and morphological evidence. Mol Biol Evol 12:679–689

Brower DL, Brower SM, Hayward DC, Ball EE (1997) Molecular evolution of integrins: Genes encoding integrin β subunits from a coral and a sponge. Proc Natl Acad Sci USA 94:9182–9187

Burke RD (1999) Invertebrate integrins: Structure, function and evolution. Int Rev Cytol 191:257–285

C. elegans Sequencing Consortium (1998) Sequence and analysis of the genome of *C. elegans.* Science 282:2012–2018

De Souza SJ, Long M, Klein RJ, Roy S, Lin S, Gilbert W (1998) Toward a resolution of the introns early/late debate: Only phase zero introns are correlated with the structure of ancient proteins. Proc Natl Acad Sci USA 95:5094–5099

Dibb NJ, Newman AJ (1989) Evidence that introns arose at protosplice sites. EMBO J 8:2015–2021

Dodemont H, Riemer D, Ledgr N, Weber K (1994) Eight genes and alternative RNA processing pathways generate an unexpectedly large diversity of cytoplasmic intermediate filament proteins in the nematode *Caenorhabditis elegans.* EMBO J 13:2625–2638

Erber A, Riemer D, Hofemeister H, Bovenschulte M, Stick R, Panopoulou G, Lehrach H, Weber K (1999) Characterization of the *Hydra* lamin and its gene: A molecular phylogeny of metazoan lamins. J Mol Evol 49:260–271

Fisher DA, Bode HR (1989) Nucleotide sequence of an actin-encoding gene from *Hydra attenuata:* Structural characteristics and evolutionary implications. Gene 84:55–64

Gendeh GS, Chung MCM, Jeyaseelan K (1997) Genomic structure of a potassium channel toxin from *Heteractis magnifica.* FEBS Lett 418:183–188

Gilbert W (1987) The exon theory of genes. Cold Spring Harbor Symp Quant Biol 52:901–905

Goldstein LSB, Gunawardena S (2000) Flying through the *Drosophila* cytoskeletal genome. J Cell Biol 150:F63–F68

Hughes AL (1992) Coevolution of the vertebrate integrin α- and β-chain genes. Mol Biol Evol 9:216–234

Hughes AL (2001) Evolution of the integrin α and β protein families. J Mol Evol 52:63–72

Humphries, MJ (1996) Integrin activation: the link between ligand binding and signal transduction. Curr Opin Cell Biol 8:632–640

Hynes RO (1992) Integrins: Versatility, modulation and signalling in cell adhesion. Cell 69:11–25

Hynes RO, Zhao Q (2000) The evolution of cell adhesion. J Cell Biol 150:F89–F95

Iacovacci S, Gagnoux-Palacios L, Zambruno G, Meneguzzi G, D'Alessio M (1997) Genomic organization of the human integrin beta4 gene. Mammal Genome 8:448–450

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921

Jiang W-M, Jenkins D, Yuan Q, Leung E, Choo KHA, Watson JD, Krissansen GW (1992) The gene organization of the human β7 subunit, the common β subunit of the leukocyte integrins HML-1 and LPAM-1. Int Immunol 4:1031–1040

Lanza F, Kieffer N, Phillips DR, Fitzgerald LA (1990) Characterization of the human platelet glycoprotein IIIa gene. J Biol Chem 265: 18098–18103

Lopez de Haro MS, Salgado LM, David CN, Boschm TC (1994) Hydra tropomyosin TROP1 is expressed in head-specific epithelial cells and is a major component of the cytoskeletal structure that anchors nematocytes. J Cell Sci 107:1403–1411

Mewes HW, Albermann K, Bahr M, Frishman D, Gleissner A, Hani J, Heumann K, Kleine K, Maierl A, Oliver SG, Pfeiffer F, Zollner A (1997) Overview of the yeast genome. Nature 387 (Suppl):7–65

Miller DJ, Hayward DC, Reece-Hoyes JS, Scholten I, Catmull J, Gehring WJ, Callaerts P, Larsen JE, Ball EE (2000) Pax gene diversity in the basal cnidarian *Acropora millepora* (Cnidaria, Anthozoa): Implications for the evolution of the Pax gene family. Proc Natl Acad Sci USA 97:4475–4480

Nyberg AM, Cronhjort MB (1992) Intron evolution: A statistical comparison of two models. J Theor Biol 157:175–190

Padgett RA, Grabowski PJ, Konarska MM, Seiler S, Sharp PA (1986) Splicing of messenger RNA precursors. Annu Rev Biochem 55: 1119–1150

Palmer JD, Logsdon JM Jr (1991) The recent origins of introns. Curr Opin Genet Dev 1:470–477

Pancer Z, Kruse M, Muller I, Muller WE (1997) On the origin of metazoan adhesion receptors: Cloning of integrin alpha subunit from the sponge *Geodia cydonium.* Mol Biol Evol 14:391–398

Ransom DG, Hens MD, DeSimone DW (1993) Integrin expression in early amphibian embryos: cDNA cloning and characterization of Xenopus beta 1, beta 2, beta 3, and beta 6 subunits. Dev Biol 160:265–275

Rezniczek GA, de Pereda JM, Reipert S, & Wiche G (1998) Linking integrin alpha6beta4-based cell adhesion to the intermediate filament cytoskeleton: direct interaction between the beta4 subunit and plectin at multiple molecular sites. J Cell Biol 141:209–225

Roy SW, Nosaka M, deSouza SJ, Gilbert W (1999) Centripetal modules and ancient introns. Gene 238:85–91

Rzhetsky A, Ayala FJ, Hsu LC, Chang C, Yoshida A (1997) Exon/intron structure of aldehyde dehydrogenase genes supports the "introns-late" theory. Proc Natl Acad Sci USA 94:6820–6825

Schuchert, P (1993) Phylogenetic analysis of the Cnidaria. Z Zool Syst Evol-Forsch 31:161–173

Spafford JD, Spencer AN, Gallin WJ (1999) Genomic organization of a voltage-gated Na+ channel in a hydrozoan jellyfish: insights into the evolution of voltage-gated Na+ channel genes. Receptors Chan 6:493–506

Stoltzfus A, Logsdon JM Jr, Palmer JD, Doolittle WF (1997) Intron "sliding" and the diversity of intron positions. Proc Natl Acad Sci USA 94:10736–10744

Stoltzfus A, Spencer DF, Zker M, Logsdon JM Jr, Doolittle WF (1994) Testing the exon theory of genes: The evidence from protein structure. Science 265:202–207

Sun H, Rodin A, Zhou Y, Dickinson DP, Harper SE, Hewett-Emmett D, Li W-H (1997) Evolution of paired domains: Isolation and se-

quencing of jellyfish and hydra *Pax* genes related to *Pax-5* and *Pax-6*. Proc Natl Acad Sci USA 94:5156–5161

Tom M, Douek J, Yankelevich I, Bosch TCG, Rinkevich B (1999) Molecular characterization of the first heat shock protein 70 from a reef coral. Biochem Biophys Res Commun 262:103–108

Van der Neut R, Krimpenfort P, Calafat J, Niessen CM, Sonnenberg A (1996) Epithelial detachment due to absence of hemidesmosomes in integrin beta 4 null mice. Nature Genet 13:366–369

Venter JC, et al. (2001) The sequence of the human genome. Science 291:1304–1351

Vibede N, Hauser F, Williamson M, Grimmelikhuijzen CJP (1998) Genomic organization of a receptor from sea anemones, structurally and evolutionarily related to glycoprotein hormone receptors from mammals. Biochem Biophys Res Commun 252:497–501

Weitzman JB, Wells CE, Wright AH, Clark PA & Law SKA (1991) The gene organisation of the human β2 integrin subunit (CD18). FEBS Lett 294:97–103

Wimmer W, Perovic S, Kruse M, Schoder HC, Krasko A, Batel R, Muller WEG (1999) Origin of the integrin-mediated signal transduction. Eur J Biochem 260:156–165

Yee GH, Hynes RO (1993) A novel, tissue-specific integrin subunit, beta nu, expressed in the midgut of *Drosophila melanogaster*. Development 118:845–858