actions and contests occur (13). In view of the wide range of nearest-neighbor distances exhibited by the population observed, it is possible that there are several ecotypes of this species-that is, geographically distinct populations with genetically determined ranges of interindividual spacing, adapted to the environments in which they occur (14).

The combination of solitary and communal behavior exhibited by this species suggests that it represents an intermediate stage in the evolution of social behavior in spiders. Perhaps the communal behavior of M. spinipes evolved as the result of an increased tolerance of conspecifics in habitats where prey was locally or seasonally abundant. In such habitats, territory size and interindividual distance could at times be reduced, and populations would be large. If web sites with proper architectural support were patchily distributed, as Agave and Opuntia are, selection might favor individuals capable of tolerating conspecifics and attaching webs together. In such situations, the advantages of group living-exploitation of habitats free of competing species, increased prey capture efficiency, architectural stability of webs, and so on-would usually outweigh the advantages gained by maintaining maximum distances from conspecifics at the cost of aggressive behavior (3, 4). Retention of interindividual spacing mechanisms like web defense, however, would ensure survival when prey availability fluctuates. The end result would be a group-living species which displays considerable variation in social grouping tendency, as M. spinipes does.

GEORGE W. UETZ THOMAS C. KANE GAIL E. STRATTON

1B

Department of Biological Sciences, University of Cincinnati, Cincinnati, Ohio 45221

References and Notes

- B. Krafit, Biol. Gabonica 2, 235 (1966); E. J. Kullmari, Am. Zool. 12, 419 (1972); R. Darchen, Insectes Soc. 20, 379 (1973); Y. D. Lubin, Zool. J. Linn. Soc. 54, 321 (1974); R. E. Buskirk, Ecology 56, 1314 (1975); J. W. Burgess, Sci. Am. 234, 100 (March 1976); V. Brach, Evolution 31, 154 (1977).
- J. 154 (1977); H. G. Fowler and J. Diehl, Bull.
 Br. Arachnol. Soc. 4, 241 (1978).
 R. R. Jackson, Rev. Arachnol. 1, 133 (1978).
 W. A. Shear, Bull. Br. Arachnol. Soc. 1, 65 (1978). (1970)
- (1978); R. E. Buskirk, in Social Insects, H. R. Hermann, Ed. (Academic Press, New York, in 4
- press), vol. 2. 5. G. W. Uetz and J. W. Burgess, *Psyche* **86**, 79 (1979).
- J. W. Burgess and P. N. Witt, Interdisc. Sci. Rev. 1, 322 (1976). 6. Ĵ
- 7. In most localities, colonies represent annual cohorts of individuals hatched at approximately the same time, or successive cohorts hatching from serial egg sacs. Mating and oviposition occur in the late fall; *M. spinipes* overwinters

(the dry season) as an egg or spiderling in the egg sac. Only in the moist tropics, where reproduc-tion occurs year-round, is there overlap of a wide variety of contemporary age classes. Marked groups of adults and juveniles showed little change during several months and usually consisted of the offspring of the residents from

- the previous year.8. J. W. Burgess, personal communication.9. Enormous colonies of web-building spiders have been observed in power lines in mountain areas in Central America, but the species was not identified because the power lines were not accessible. Near Cordóba, colonies had spilled over onto the ground and were identified as those of *M. spinipes*. We estimated colony size in several large groups by measuring the volume of the communal web and then extrapolating from density counts. Several of the larger groups contained an estimated 5800 to 6500 individuals
- 10. Prey availability was estimated by observation of the number of insects approaching a 1-m^3 colony web for a specified time period and then correction of this estimate by the number of

- hours of favorable weather conditions in a day. S. E. Riechert, Symp. Zool. Soc. London 42, 211 (1978). 11.
- 12
- 211 (19/8).
 Am. Nat. 117, 871 (1981).
 R. E. Buskirk, Anim. Behav. 23, 560 (1975).
 It is also possible that the several populations observed represent distinct species or subspecies. Although the genitalia of all specimens match the drawings in the species description of F. O. Pickard-Cambridge [Biol. Cent. Am. 2, 457 (1903)] there are considerable difference in the second seco 14. 457 (1903)], there are considerable differences in body size, leg length, and life history, as well as social grouping tendency. Metepeira spinipes is the only group living spider known in the genus Metepeira [H. W. Levi, Bull. Mus. Comp. Zool. Harv. Univ. 148, 185 (1977)], although T. W. Schoener (personal communication) has found groups of *M. daytoni.*
- We thank R. Hollis and M. K. Uetz for assist-ance in the field. Supported by the National 15. Geographic Society, the American Philosophi-cal Society, and the University of Cincinnati Research Council.

29 January 1982; revised 13 April 1982

Diverse Mechanisms in the Generation of Human β-Tubulin Pseudogenes

Abstract. The sequence of two human β -tubulin pseudogenes is described. One contains an intervening sequence but lacks sequences encoding the 55 N-terminal amino acids of the polypeptide chain. A second has no introns but has a polyadenylate signal and an oligoadenylate tract at its 3' end, and it is flanked by a short direct repeat. These sequences have arisen by different mechanisms, including one that probably involves reverse transcription of a processed messenger RNA and reintegration of the complementary DNA copy into the genome.

With few exceptions, all expressed eukaryotic genes contain intervening sequences (introns), regions of DNA that interrupt the coding sequence and that are spliced out of the primary gene transcript as part of the generation of cytoplasmic messenger RNA (mRNA). Any of a number of genetic events including

Sacl

deletion, insertion, or acquisition of one or more stop codons within the coding sequence can lead to a failure to yield a functional transcript. Such a gene is therefore termed a pseudogene. Among the pseudogenes thus far described, a curious form is that in which the intervening sequences present in the ex-



Eco R I

Smai

Fig. 1. Sequencing strategies. A 4.1-kb Hind III-Eco RI restriction fragment (11B) and a 1.55-kb Eco RI fragment (1β) , each containing B-tubulin sequences, were isolated from the appro-Charon priate 4Å lambda clones (3) by electrophoresis in low melting temperature agarose and extraction with phenol. A further 1.65-kb Bam HI-Hind III fragment containing B-tubulin (1B) was subcloned into pBR322. Restriction endonuclease digestion of these DNA's was performed with the enzymes indicated and the fragments ligated into appropriately

cleaved replicative forms of bacteriophage M13 mp8 and mp9. In some cases, plaques containing the DNA to be sequenced were identified on nitrocellulose replicas (16) by hybridization with a nick-translated chicken β -tubulin cDNA probe (17).

pressed homolog have been precisely eliminated (1, 2). Because the mechanism whereby intervening sequences are removed involves splicing at the RNA level, it is logical to ascribe the generation of intron-deficient pseudogenes to a process of reverse transcription, in which a spliced RNA transcript is copied into DNA and the DNA copy is reinserted into the genome. Here we present evidence that two genomic human β tubulin sequences lack introns but contain 3-polyadenylate [poly (A)] tracts and that each is derived independently by reverse transcription of different spliced mRNA's. A third intron-containing pseudogene which has lost a substantial proportion of its 5' coding sequences, probably as a consequence of deletion, is also described.

Alpha- and beta-tubulins are the principal proteins of microtubules, filamentous structures found in association with several subcellular structures in eukaryotic cells, including the mitotic apparatus, cilia, flagella, and elements of the cytoskeleton. Evidence based on amino acid sequencing, immunological crossreactivity, and the ability of cloned tubulin-specific complementary DNA (cDNA) probes to hybridize to DNA fragments from diverse species under stringent conditions points to a close evolutionary conservation among tubulins. In humans, the genes that encode these proteins each constitute a multigene family consisting of about 15 to 20 members, several of which have been isolated as recombinant fragments (3-5). These sequences include examples that are short, that is, approximately the size required to encode a human tubulin mRNA. The complete sequence of two such genes was determined by the stratagem depicted in Fig. 1. In Fig. 2, the sequence data is compared with the sequence of a second short human β -tubulin gene (6) and a cloned chicken β - tubulin cDNA probe (7). Two of the human sequences are devoid of introns, and all three contain lesions (in-frame termination codons and deletions or insertions that cause frameshifts and resultant premature termination signals) that preclude the generation of a functional product. Like 46 β , the sequence of 11 β includes at its 3' end a poly(A) signal (AATAAA) and an oligo(A) tract 16 base pairs further downstream. In addition, 11 β is flanked by a short direct repeat of sixteen base pairs that differs completely from the direct repeat flanking 46 β .

The sequence of 1β reveals two singlebase deletions (in the first codon position of glycines 269 and 400) that lead quickly to in-phase termination codons at amino acid positions 361 and 405. However, in contrast to 46 β and 11 β , this gene contains a short (120 base pairs) intervening sequence that interrupts the coding sequence at glycine 94 and is flanked by correct consensus splice signals. There

pT2 18 118 468	TAAGGCCCGGGAGCGGGGCGCGGCCGGCCGGCAGGGCGCGGGCGCGGGCGCGGGCGCGGGCGCGGGCGCGGCGC															CGGA AGGA CACG FTAT															
pT2 18 118 468	GAG TCA AGC TTT	CGGA TTTC CAGC CTTC	GGT CAA CCAA CCCA	GACG FATG GCCC CGTA	GAGO AAGO GCC CATA	CGGC GAAC IGC ACC	GAGO GAGI CCG1 TTGI	CAGCO AATC' FCCGI AAGTO	GCGG TCTG AGCC GAGC	CACCO FCACS SCCTO AAAA	GCA GGC GCCA AAAT	GACA AATA GACA FAAA	CCGGC FCACI CGCCC FTTTI	CATC AGAGO CATG AACC	Met ATG CAGA(Arg CGT CTGA A G A G	Glu GAG TGGA T A	Ile ATC GATTC G	Val GTG GTTT C	His CAC GTGG0	Ile ATC GTAT T	Gln CAG TCTA	Ala GCC TAGA T	10 Gly GGC TTTG T	Gln CAG TGGG	Cys TGC ATTG C T	Gly GGC TTGT	Asn AAC rGGGG	Gln CAG TĂA/ T	Ile ATC AAAA	Gly GGC TATG G T
рТ2 1В 11В 46В	Ala GCT CATC C	Lys AAC GTTA	20 Ph TT GAT	e Tr C TG ACTC	p G G G AAA	lu ' AG (FCT(A	Val GTC CGAC	Ile ATC CTA	Ser AGC CCAA T	Asp GAT GGGC	Glu GAG CCCT A	His CAC TTCT/ T T	Gly GGC AGAAS	30 Ile ATC FATCC A	Asp GAT CATG C C	Pro CCC ATTT	Thr ACC FTTT G T	Gly GGC TGGAG	Ser AGC CCAG A C	Tyr TAC FATC	HİS CAC ACAA GTG	Gly GGG AGTT(Asp GAC CTAT A	40 Ser AGC TTTG TCG	Asp GAC ATAA	Leu CTG AACA? T	Gln CAG TTAAG G	Leu CTG CTTT	Asp Gag Faga <i>i</i> C	Arg AGG AAAAC CA C C	Ile ATC CAAG
PT2 18 118 468	Asn AAC TAGO G (2 TC	Val GTC GCTC	50 Ty: TA ACT	r Ty C TA FTTT	r As C Al CCTI	sn (AT (ATT C 3	Gly GAG TTTC A	Ala GCC TA	Thr ACA T T T	Gly GGT TC A	Asn AAC GG C T GG	Lys AAG A	Tyr TAC T T T	60 Val GTC G G T	Pro CCC T T	Arg CGT A G	Ala GCC A	Ile ATC G A	Leu CTG T * C	Val GTT G C G	Asp GAC T	Leu CTG A	Gly GAG A	70 Pro CCC T T	Gly GGC G G G G	Thr ACG C C	Met A T G	Asp GAC	Ser TCG AGC AGT T	Val GTG A T C T	Arg CGC A G
pT2 18 118 46B	Ser TCC T G A	Gly GGC A	80 Pr CC AA	o Ph C TT A F F	e G T G A	ly (GA (C	Gln CAG GCT T	Ile ATC C T T	Phe TTC T T	Arg CCA A AGG AG	Pro CCC T A	Asp GAC	Asn AAC GT T	90 Phe TTT A	Val GTC A A	Phe TTT CA	Gly GGT	ATGT!	[TTT	CCAG	∖ AGG'	TTCC/	ACCA	GGAG	GAGG	GGGG	GATG	CTTT	ACTGO	TGC	сстт
pT2 18 118 468	СТС	ŤŦŦĨ	CAC	CTTT	CTT	cco	CTG	CTGG'	r T TC'	FCTT	FTT G(GCCAG	CAGG	Gln CAG FA C	Ser AGC TCT T TCT	Gly GGG	Ala GCC T A	Gly GGC T	Asn AAC T	100 Asn AAC	Trp TGG	Ala GCC	Lys AAG A G A	Gly GGG C T C	His CAC	Tyr TAC	Thr ACG A	Gly GAA G G G G	Gly GGC A T	110 Ala GCT C G C	Glu GAG C
pT2 1B 11B 46B	Leu CTG G	Val GTC A C	As GA	p Se C TC G AA F F	r Va T G' C C	al I IG C C C	Leu CTG A	Asp GAT G	Val GTG	120 Val GTG CG A	Arg AGG A C	Lys AAG C C	Glu GAG ***	Ser TCG AGT GT G A	Glu GAA G G G	Ser AGC A	Cys TGC T T	Asp GAC	Cys TGC G T	130 Leu CTG	Gln CAG T	Gly GGC T	Phe TTC	Gln CAG	Leu TTG A C C C C	Thr ACC GT	HİS CAC T	Ser TCG C A	Leu CŤG	140 Gly GGC T	Gly GGC G G G
рТ2 18 118 46В	Gly CGC	Thi ACC P P	Gl GGG A	y Se C TC	r G 3 G C T	ly I 3G J T A	Met ATG	Gly GGC	Thr ACC T G T	150 Leu CTC G G	Leu CTC T	ile ATC G	Ser AGC A	Lys AAG	Ile ATC T	Arg CGC A A AT A	Gly GAG A	Gly GAG A	Туг ТАС Т	160 Pr o CCC G T	Asp GAC A T	Arg CGC G T	Ile ATC	Met ATG	Asn AAC T T	Thr ACG T C C	Phe TTC	Ser AGC T	Val GTC A G	170 Met ATG G G	Pro CCC T T
	550																										S	CIEN	CE.	VOL.	217

рТ2 18 118 468	Ser TCC T G A	Pro CCC	Lys AAG A	Val GTG	Ser TCG * T	Asp GAC ***	Thr ACG T C	Val GTG	180 Val GTG A C	Gly GAG	Pro CCC T	Tyr TAC	Asn AAT C C	Ala GCC G C T	Thr ACC GTT TG	Leu CTT G G G	Ser TCT C C	Val GTG A C A C	190 His CAC T	Gln CAG	Leu CTG T	Val GTG A T A	Gly GAG	Asn AAC T T	Thr ACG G A A T	Asp GAC T T T	Glu GAG CC	Thr ACC TGT	200 Tyr TAC T	Cys TGC
рТ2 18 118 468	Ile ATC T T	Asp GAC A	Asn AAC T G	Gly GAG	Ala GCC G G	Leu CTG C C C	Tyr TAC T G T	Asp GAC T	210 Ile ATC	Cys TGC	Phe TTC G	Arg CGC T A	Thr ACC T	Leu CTG C	Lys AAG G G	Leu CTC G G G G G	Thr ACC G	Thr ACT A G A	220 Pro CCC	Thr ACG C C C	Tyr TAC T T	Gly GGG	Asp GAC T T	Leu CTC AC	Asn AAC G	His CAC T	Leu CTG A C	Val GTG ACA	230 Ser TCG C T AA	Asp GCC TTG A
pT2 18 118 468	Thr ACC	Met ATG	Ser AGC GAG	Gly GGC A G T T	Val GTG A A A C C	Thr ACC T	Thr ACC	Cys TGC C	240 Leu CTT C C	Arg CGC G	Phe TTC	Pro CCC G G T	Gly GGC T	Gln CAG	Leu CTG C C C	Asn AAC T T	Ala GCC A G T	Asp GAC	250 Leu CTG C	Arg CGC A	Lys AAG	Leu CTG T	Ala GCG T C	Val GTC G G T	Asn AAC	Met ATG	Val GTG C	Pro CCT C C C	260 Phe TTC	Pro CCC A
рТ2 1в 11в 46в	Arg CGG C T C T C	Leu CTG C	His CAC T	Phe TTC	Phe TTC T	Met ATG	Pro CCG C A T	Gly GGC *	270 Phe TTC T	Ala GCC ***	Pro CCG A *** T	Leu CTG C A C	Thr ACG ** C	Ser AGC G	Arg CGC AG C AT	Gly GGC A	Ser AGC	Gln CAG	280 Gln CAG T	Tyr TAC T	Arg CGA G A	Ala GCC AG	Leu CTG C	Thr ACG T C C A	Val GTG	Pro CCC G G	Glu GAG C	Leu CTG C C C	290 Thr ACG C C	Gln CAG CT
рТ2 18 118 46В	Gln CAG	Met ATG G C	Phe TTC T	Asp GAC T T	Ser TCC G G G	Lys AAG CGC	Asn AAC T G	Met ATG CC	300 Met ATG	Ala GCC T T	Ala GCC	Cys TCC GT CG GT	Asp GAC T	Pro CCC T GG	Arg CGC A G	His CAC GT	Gly GGC	Arg CGC T A	310 Tyr TAC	Leu CTG C	Thr ACG A G A C	Val GTG	Ala GCT C C	Ala GCC TG A T	Ile ATC T G G GA	Phe TTC	Arg CGA G G T	Gly GGC T	320 Arg CGC AAG T G	Met ATG C
PT2 18 118 468	Ser TCC	Met ATG CC	Lys AAG	Glu GAG A	Val GTG C	Asp GAC T	Glu GAG C C	Gln CAG A	330 Met ATG C	Leu CTG C	Asn AAC TC TC	Val GTG A C	Gln CAG	Asn AAC C G	Lys AAG G	Asn AAC	Ser AGC	Ser AGC	340 Tyr TAC G	Phe TTT C	Val GTG	Glu GAG A	Trp TGG	Ile ATC T	Pro CCC	Asn AAC	Asn AAC T	Val GTG C A A C	350 Lys AAG	Thr ACG GT GT A
pT2 18 118 468	Ala GCC T A T	Val GTC	Cys TGC T	Asp GAC	Ile ATC	Pro CCC G A A	Pro CCG C C T	Arg CGT G CC	360 Gly GGC G	Leu CTC G	Lys AAG _GC	Met ATG	Ser TCC G T G A	Ala GCC T T	Thr ACC	Phe TTC C	Ile ATC T	Gly GGC A	370 Asn AAC GT	Ser AGC A	Thr ACG	Ala GCC G	Ile ATC	Gln CAG A	Glu GAG	Leu CTC A G	Phe TTC T T	Lys AAG T G *	380 Arg AGG CAC	Ile ATC G
pT2 1B 11B 46B	Ser TCG T A	Glu GAG C	Gln CAG T	Phe TTC	Thr ACG T A T	Ala GCC A	Met ATG	Phe TTC	390 Arg CGG AAA A C	Arg CGC A G A G	Lys AAG A	Ala GCC T	Phe TTC T	Leu TTG G C A C C	HİS CAC	Trp TGG	Tyr TAC	Thr ACC A TG A	400 Gly GGC *	Glu GAG A A	Gly GGC G	Met ATG	Asp GAT C C C	Glu GAG ATA	Met ATG AC	Glu GAG A	Phe TTC T A	Thr ACC GGG	410 Glu GAG A	Ala GCG T C T
pT2 1B 11B 46B	Glu GAG A A C	Ser AGC AT	Asn AAC	Met ATG C	Asn AAC C T T	Asp GAC T	Leu CTG T C	Val GTC A G	420 Ser TCA C C T	Glu GAG	Tyr TAC T	Gln CAG A	Gln CAA G G	Tyr TAC TTT	Gln CAG F A	Asp GAT C	Ala GCC T	Thr ACC AA TG	430 Ala GCT A C A	ASP GAT T C G A	Gln CAG TA G G	Gln CAG G G G	Gly GGT AA AG	Glu GAA T G T	ATG	Phe TTT *** C G	Glu GAA C GT	Glu GAG T	Glu GAA TC G G	440 Gly GGA ACG AG CC
pT2 18 118 468	Glu GAG A A	Glu GAG A	Asp GAT CA TTG ***	Glu GAG A	Ala GCA ** C C	OP TGA G G A	AGTO AGCO CCCO AGAO	GAA CAGA CAAG CAAG CAAG CAAGC	ATGGO AGATA IGAAG CCCCA	GTAGO AAGGO GCTCI AATCI	GAGTO GACAJ FTGCA AACCJ	CAAGI ITAAC AGCTC ICGAC	IGTAC CTGTC GGAG GGCT	GTCTO GAGAO FGAGO FCTC <i>I</i>	GAGCI GAAGO GGGCI AGTTO	AGGC/ CTGTC AGGTC CCCT1	AAGTO GCCGO GGCGO FAGCO	GTTC/ CGGA(CCGG? CGTC?	ATTG/ STCG(FCCA/ FTGA(AGAGO CTTA AGGCO CTCA	GAGA CAGA CAGC ACTG	ATCTO ACAGI AGTGI CCCGI	TGC TTTC: CTG TTTC:	AGTTO FCAT ACCTO CTCTO	GTGCI FAGAI CCAGA CCCTC	GAAG GAGI GCCZ GCCZ	CATO GTTI TCTI TCTI	CGTI CTCC GCTG TGTI	TCTI TGCA TCGA	'AAT \GCA \CGC IGCC
рТ2 1В 11В 46В	TTCC CTCC TGTC TCT/	GTGC CAAAI CCCCI ATCT	rctco Accci Agcti rgcti	CGCTC ACTC PTCCC PTTTC	GTTG(IGCA(CCCA(GTTT)	CTTT CTGCA CCAGO FTTC	FGTC AGCAC CTTG7 FTCTC	AGCAC CAGTO FCAC GGGG	GTTTT GAATO FCACO AGGGT	IGTA1 GATA1 GCTAC ICTAC	FCCTO FGCAO GGGC1 GAACZ	CGACI CTCAC FCCCI ATGCC	rgtc(CCAT rtgc(Ctgg(CGATO PAGCI CACCI CACAI	GTAAG FTCG/ FCCTG FAGT/	CAGT ACACI GCAG AGGCI	IGCA AGGG IGTT ACTC <u>I</u>	AAAA ACTGI FACAG AATAI	TACT AGGG CCGT AATA	FCAG AGAC CCTC CTTG	AGTC' AGGT(CCCA(TTTG'	TTCTC GGGG CCTAC TTGG	GTTG AGCAC GGCC/ AAAA/	AAATG GCTG AGTG AAAA	GTTA ACAGO IGTGO AAAA?	ACTI CATI TGCI		AACA ITCTT ITCTC IGAGC	TAAA GCTG TGTC TGTC	IGGC IACA ITTA IAGG
рТ2 1в 11в 46в	TTT TCT/ TTG ATC	TTGT(ACTA) CAGC' CCTT(GTTC ACCT ICCA GAGC	T <u>AAA</u> IGAAC GGCC CTGGC	AAAA JAGT IGAC JAGG	AAAA FTGA ATTT FCAAC	AAAA IGTTO IATGO GGCTO	AA CAGTO GATTO GCAG	GCATA GTT1	\CTT# [TTT#	ATTA AGCTO	ACTT! GTT!	AAA) IGTGI	AAAT) ITTA:	AGCAJ FATT	AATT: FTCAC	TATTO GGAO	STAAI CTAC'	AGTG(TT <u>AA</u>	GATC(CCTT CTCT	IGTT: ATTG(CAA/ CCTG	AGTG' FCAG	FTTGC AT <u>AA</u> J	CAGO	CATO	CAGP	CTAC	agt Iaga

1B GTGGATTTGCAGGGAGCCCACT 11B CTGACCACAAAGAAATCCAAA

Fig. 2. Sequence of three human β -tubulin pseudogenes. The sequences of 1 β and 11 β were determined by the dideoxy chain terminator method (18) and are compared with the sequence of 46 β (6). Asterisks denote deleted bases; circles denote inserted bases. Both have been introduced so as to retain maximum homology (represented by gaps within the coding regions) with the prototype chicken β -tubulin cDNA sequence pT2 (7). Termination codons are boxed. The poly(A) signals and the oligo(A) tracts are underlined. Flanking direct repeats are doubly underlined. Abbreviations: Ala, alanine; Arg, arginine; Asp, aspartic acid; Asn, asparagine; Cys, cysteine; Gln, glutamine; Glu, glutamic acid; Gly, glycine; His, histidine; Ile, isoleucine; Leu, leucine; Lys, lysine, Met, methionine; Phe, phenylalanine; Pro, proline; Ser, serine; Thr, threonine; Trp, tryptophan; Tyr, tyrosine; and Val, valine.

6 AUGUST 1982

is also substantial divergence from the other sequences toward the 3' end, and 1B contains no discernible poly(A) addition site or significant oligo(A) tract within 700 base pairs further downstream; nor is there evidence of direct or inverted repeat sequences in the regions immediately flanking this gene. Toward the 5' end (at amino acid 55) an unexpected feature is that all homology, both with respect to the chicken cDNA probe and the intron-lacking genomic sequences, is abruptly lost. Inspection of the region spanning the threonine 55 codon reveals a sequence (TTTCTACAGGT) (T, thymine; C, cytosine; G, guanine; A, adenine) that closely resembles a consensus acceptor splice signal, including the obligatory dinucleotide AG. Indeed, the structure of a 6.8-kb (kilobase) human βtubulin gene, determined by heteroduplex analysis with the cloned chicken β tubulin cDNA, shows the presence of a short intervening sequence at approximately this location (3). However, Southern's hybridization experiments (8) on overlapping recombinant phage that contain the 1B pseudogene and extend 15 kb upstream from the region sequenced show no further evidence of hybridization with human or chicken β -tubulin probes containing 5' coding sequences. Thus, although the existence of an intervening sequence greater than 15 kb cannot be ruled out, it seems more likely that some DNA rearrangement (deletion, inversion, recombination, or transposition) with a breakpoint upstream from amino acid 55 resulted in a loss of sequences encoding the 5' end of this gene.

An additional feature of 1β is that the largest (4.9 kb) intron contained in the 6.8-kb gene (3) is absent. One possibility is that this intervening sequence was eliminated by reverse transcription of an RNA processing intermediate, and the cDNA copy reintegrated into the genome. This mechanism would appear unlikely, however, in view of the absence of a poly(A) signal, poly(A) tract, or flanking direct repeats. A more plausible explanation is that 1ß was generated solely by mutational events from a functional B-tubulin gene that lacks the 4.9kb intron. In contrast to the known intron-containing pseudogenes of the globin gene families (9), 1 β does not appear to be closely linked to other β-tubulin sequences. A transposition event may have conveyed 1B to its present location, or it may represent a member of the Btubulin multigene family that was already dispersed, and subsequently mutated so as to be no longer functional.

The addition of A residues to the 3' ends of mRNA molecules is a post-transcriptional event. Hence, the occurrence of 3'-poly(A) tracts in the two intronlacking human \beta-tubulin pseudogenes (Fig. 2) is suggestive of a mechanism involving the reverse transcription of processed mRNA. How could such molecules be generated and, once generated, how could they become integrated into the genome? Among the requirements would be a reverse transcriptase and a functional primer for the synthesis of plus and minus strands. Although no obvious molecules exist that might serve as primers, the rarity of successful synthesis could be a consequence of improbable events. A possible integration mechanism might then involve recombination with a transposable element, for example, retrovirus DNA. Indeed, recombination of retroviral sequences with a fragment of genomic DNA and loss of intervening sequences during the RNA phase of the virus life cycle-a process known to occur with incorporated host genes (10)—can satisfactorily explain the observation that two intron-lacking mouse α -globin pseudogenes both retain homology with respect to functionally expressed genes 5' to the mRNA capping site (1, 2). This mechanism is also consistent with the detection of flanking retrovirus-like elements (11) and the observation that at least one of the pseudogenes is dispersed relative to its expressed homolog (12). However, neither of these pseudogenes contains a 3'-oligo(A) tract and, although a spliced mRNA or its cDNA cognate could conceivably recombine directly with a retrovirus sequence, a much simpler mechanism would involve the direct integration of a cDNA copy into a staggered host chromosomal break. An aberrant human immunoglobulin λ light chain transcript (13) may also have been generated in this fashion. Direct repeats could then arise as a consequence of DNA repair at the integration site (14).

The sequences corresponding to the 3' untranslated and 5' untranslated regions of the two intron-lacking human B-tubulin pseudogenes are of different length and show no significant homology. It seems likely, therefore, that 46B and 11B arose by reverse transcription of two independently expressed β-tubulin genes. If we assume that reintegration of cDNA copies can occur anywhere in the genome, the propagation of intron-lacking pseudogenes in the germ line would require that there be no interference with the expression of other essential genes. The accumulation of a number of mutations in these integrated cDNA's suggests that they were probably nonfunctional. Eventually, such progressive changes might result in a region of DNA that is unrecognizable with respect to the original transcript. That such a degenerative process indeed occurs is suggested by the greater number of insertions, deletions, and amino acid substitutions (compared with the functional chicken cDNA sequence) in 11β , and implies that the integration of this reverse transcript significantly predates the integration of 46β. The sequence of a third intronlacking human B-tubulin pseudogene that also contains multiple point mutations, small deletions, and insertions throughout its length is consistent with this conclusion (15). It remains to be seen to what extent the genomes of higher vertebrates are populated by dispersed reintegrated cDNA molecules generated by reverse transcription of spliced mRNA's. In any event, it is clear that several distinct and independent pathways can lead to the existence of homologous but apparently functionless genelike sequences.

> C. D. WILDE C. E. CROWTHER N. J. COWAN*

Department of Biochemical Sciences, Princeton University, Princeton, New Jersey 08544

References and Notes

- Y. Nishioka, A. Leder, P. Leder, *Proc. Natl.* Acad. Sci. U.S.A. 77, 2806 (1980).
 E. F. Vanin, G. I. Goldberg, P. W. Tucker, O. Smithies, *Nature (London)* 286, 222 (1980).
 N. J. Cowan, C. D. Wilde, L. T. Chow, F. C. Wefald, *Proc. Natl. Acad. Sci. U.S.A.* 78, 4877 (1981) (1981)

- (1981).
 C. D. Wilde, L. T. Chow, F. C. Wefald, N. J. Cowan, *ibid*. **79**, 96 (1982).
 C. D. Wilde, C. E. Crowther, N. J. Cowan, J. Mol. Biol. **155**, 533 (1982).
 C. D. Wilde, C. E. Crowther, T. P. Cripe, M. G.-S. Lee, N. J. Cowan, Nature (London) **297**, 83 (1982).
- G.-S. Lee, N. J. Cowan, Nature (London) 297, 83 (1982).
 P. Valenzuela, M. Quiroba, J. Zaldivar, W. J. Rutter, M. W. Kirschner, D. W. Cleveland, *ibid.* 289, 650 (1981).
 E. Southern, J. Mol. Biol. 98, 503 (1975).
 P. F. R. Little, Cell 28, 683 (1982).
 S. P. Goff, E. Gilboa, O. N. Witte, D. Baltimore, *ibid.* 22, 777 (1980).
 K. Luchern, V. Leitze, P. Leder, F. Kuff, Na.

- K. Lueders, A. Leder, P. Leder, E. Kuff, *Nature (London)* 295, 426 (1982). 11.

- ture (London) 295, 426 (1982).
 12. A. Leder, D. Swan, F. Ruddle, P. D'Eustachio, P. Leder, *ibid*, 293, 196 (1981).
 13. G. F. Hollis, P. A. Hieter, O. W. McBride, D. Swan, P. Leder, *ibid*, 296, 321 (1982).
 14. S. W. Van Arsdell, R. A. Denison, L. B. Bern-stein, A. M. Weiner, T. Manser, R. F. Geste-land, Cell 26, 11 (1981).
 15. T. P. Cripe and N. J. Cowan, unpublished data.
 16. W. D. Benton and R. W. Davis, Science 196, 180 (1977).
- 180 (197
- D. W. Cleveland, M. A. Lopata, R. J. McDon-ald, N. J. Cowan, W. J. Rutter, M. W. Kirschner, *Cell* **20**, 95 (1980).
 F. Sanger, A. R. Coulson, B. G. Barrell, A. J. H. Smith, B. Roe, *J. Mol. Biol.* **143**, 161 (1980).
 Supported by NIH grant GM26456. C.D.W. ocknowledges the award of a Muscular Dystro-
- acknowledges the award of a Muscular Dystrophy postdoctoral fellowship. Present address: Department of Biochemistry,
- New York University Medical Center, New York 10016.

9 April 1982; revised 24 May 1982