Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 8507–8511, August 1995
Evolution

Seven newly discovered intron positions in the triose-phosphate isomerase gene: Evidence for the introns-late theory

(molecular evolution/exons/intron gain/intron sliding)

JOHN M. LOGSDON, JR.*, MICHAEL G. TYSHENKO†, COLLEEN DIXON*‡, JONATHON D.-JAFARI*§, VIRGINIA K. WALKER†, AND JEFFREY D. PALMER*¶

*Department of Biology, Indiana University, Bloomington, IN 47405; and †Department of Biology, Queen's University, Kingston, ON Canada K7L 3N6

Communicated by Charles S. Levings III, North Carolina State University, Raleigh, NC, June 9, 1995

ABSTRACT **The gene encoding the glycolytic enzyme triose-phosphate isomerase (TPI; EC 5.3.1.1) has been central to the long-standing controversy on the origin and evolutionary significance of spliceosomal introns by virtue of its pivotal support for the introns-early view, or exon theory of genes. Putative correlations between intron positions and TPI protein structure have led to the conjecture that the gene was assembled by exon shuffling, and five TPI intron positions are old by the criterion of being conserved between animals and plants. We have sequenced TPI genes from three diverse eukaryotes—the basidiomycete *Coprinus cinereus*, the nematode *Caenorhabditis elegans*, and the insect *Heliothis virescens*—and have found introns at seven novel positions that disrupt previously recognized gene/protein structure correlations. The set of 21 TPI introns now known is consistent with a random model of intron insertion. Twelve of the 21 TPI introns appear to be of recent origin since each is present in but a single examined species. These results, together with their implication that as more TPI genes are sequenced more intron positions will be found, render TPI untenable as a paradigm for the introns-early theory and, instead, support the introns-late view that spliceosomal introns have been inserted into preexisting genes during eukaryotic evolution.**

The surprising discovery of spliceosomal introns in 1977 was soon followed by considerable speculation about their origin, evolution, and significance (1–3). Nearly 20 years later, the issue is very much alive and has long since become polarized into two opposing theories. The introns-early theory, or exon theory of genes, posits the presence of many introns in the common ancestor of all life, followed by massive, often complete, intron loss in many independent lineages (4, 5). Introns are thought to have functioned in the primordial assembly of protein genes by promoting the recombinational shuffling of short exons, each encoding 15–20 amino acid units of protein structure (6–8). The other theory, termed introns-late, posits that spliceosomal introns were not present in the common ancestor of life but, instead, arose and spread within eukaryotic evolution (9–11); therefore, these introns could not have played any role in ancient gene and protein assembly.

A major part of the evidence in favor of the introns-early theory has been supplied by the ancient gene encoding the glycolytic enzyme triose-phosphate isomerase (TPI; EC 5.3.1.1; refs. 6–8 and 12). As soon as the first eukaryotic TPI gene was sequenced, a correspondence was noted between exons and secondary structural elements, with all six chicken introns falling at or near the ends of α-helices and β-strands (13). More TPI intron positions were discovered in 1986 from a plant and a fungus (6). Five introns are located in the same positions in plant and animal TPI genes, indicating that these introns were in place prior to the presumably ancient divergence of these taxa (6). Since these new data did not support a straightforward correlation between exons and secondary structural elements, it was instead suggested that the presumed ancestral exons in TPI encoded compact polypeptide "modules" (6, 7, 14). Eleven hypothetical exons were proposed based on the sum total of introns then known in animal, plant, and fungal genes. One of these exons did not represent a compact module (6, 7, 14), leading Gilbert *et al.* (6) to predict "that in some other [TPI] gene there will be an intron that breaks up this exon." An intron position that fulfilled this introns-early prediction was recently reported in a mosquito, *Culex tarsalis* (15). However, as noted by Doolittle and Stoltzfus (16), for this prediction to have been robust, most, if not all, other possible intron positions would have to have been forbidden. Statistical evaluations of exon/protein structure correlations in TPI have produced seemingly conflicting results. The multiple analyses of Stoltzfus *et al.* (17) cast serious doubt on the supposed correlations of TPI exons and protein structure, whereas a significant correspondence was found in the single analysis of Gilbert and Glynias (8).

The discovery of the *Culex* TPI intron (15) prompted our laboratories to sequence more TPI genes from diverse eukaryotes. The rationale of the Ontario authors was that if the *Culex* intron were ancient, as predicted by the exon theory of genes (6–8), it should be found in other insects; thus, the lepidopteran *Heliothis virescens* was chosen for examination. The opposite reasoning, taken by the Indiana authors, was that the *Culex* intron represented a recent insertion whose position fortuitously fit Gilbert *et al.*'s prediction (6); we therefore predicted that as more TPI genes were sequenced, more intron positions would be discovered and that these sites would fall randomly with respect to TPI protein structural elements. To search for additional TPI introns, we chose the basidiomycete *Coprinus cinereus* and the nematode *Caenorhabditis elegans*, which represent previously unsampled groups that contain many introns not present in homologous genes from other fungal or animal lineages (J.M.L. and J.D.P., unpublished data). The TPI genes that we have sequenced contain a total of seven novel intron positions, all of which are unanticipated by and inconsistent with the exon theory of genes; instead, these data provide strong support for the introns-late theory of eukaryotic gene evolution.‖

MATERIALS AND METHODS

Two overlapping TPI fragments were generated from *Heliothis* genomic DNA using primer pairs corresponding to amino

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: TPI, triose-phosphate isomerase.

‡Present address: Eli Lilly and Company, Inc., Lilly Corporate Center, Indianapolis, IN 46285.
§Present address: Genetics Therapy, Inc., 938 Clopper Road, Gaithersburg, MD 20878.
¶To whom reprint requests should be addressed.

‖The sequences reported in this paper have been deposited in the GenBank data base (accession nos. U23079–U23081).

acids 8–13 and 168–173 and 121–127 and 228–233 of *Gallus* TPI. PCR fragments were cloned and sequenced at least twice per strand on an automated DNA sequencer. Intron positions were inferred by inspection of the genomic sequence using the conserved nucleotide sequence encoding TPI and the presence of canonical spliceosomal intron splice sites. A TPI cDNA clone (cm13b12) from *Caenorhabditis* (from R. Waterston, Washington University, St. Louis) was used to screen a set of *Caenorhabditis* genomic λ clones (from A. Coulson, Medical Research Council Laboratory, Cambridge, U.K.) that mapped near the same genomic region as cm13b12. Both strands of the TPI region from λ clone VT#YL77 were then sequenced manually. Intron positions were verified by partially sequencing cDNA clone cm13b12. *Coprinus* RNA (from M. Zolan, Indiana University, Bloomington) was reverse-transcribed and used for PCR amplification to generate a TPI fragment used to screen a *Coprinus* cosmid library (from M. Zolan). The TPI region from one cosmid was sequenced manually. A full-length TPI cDNA clone was sequenced in order to verify *Coprinus* intron positions. Analyses of protein structure correlations and generation of random intron positions were carried out using the ABACUS program (ref. 17; A. Stoltzfus, D. F. Spencer, and W. F. Doolittle, personal communication) on a Sun Sparcstation. The reference protein was the 247-residue chicken muscle TPI enzyme (Brookhaven Protein Data Bank entry pdb1tim).

RESULTS AND DISCUSSION

Two complete TPI gene sequences were determined, from the nematode *Caenorhabditis elegans* and the basidiomycete *Coprinus cinereus*, along with an 86% complete sequence from the insect *Heliothis virescens*. The *Heliothis* region does not contain an intron at the *Culex* position but does contain three introns known only in this species, plus a fourth intron shared by other animals (Fig. 1). In addition, three of the five introns in the *Coprinus* TPI gene are located at novel positions, as is one of the two *Caenorhabditis* introns. Including these seven introns, the total number of intron positions that are unambiguously distinct with respect to the TPI alignment is now 21 (Fig. 1). These results undermine the introns-early theory (4–8, 15) in four significant ways.

More Introns Continue to Be Found in TPI. These data clearly show that the *Culex* intron (15) was not the last to be found in TPI; equally clearly, they imply that as more TPI genes are sequenced, even more intron positions will be found, rendering the gene too fractured to fit the exon theory of genes (6–8). This is already the case for other genes; for example, in the more widely sequenced gene for glyceraldehyde-3-phosphate dehydrogenase (GAPDH) there are now 47 known intron positions (in a 333-codon gene; ref. 18). Therefore, the average exon would be only 7.2 codons in a hypothetical (according to the introns-early theory) ancestral GAPDH gene containing all known introns (19). This is significantly smaller than the 15- to 20-codon exons postulated by the exon theory of genes (6–8), and many of these exons are so small as to be of doubtful utility in ancient gene assembly (19). With 21 known introns in TPI, the average size for inferred ancestral exons is 11.2 codons and can only continue to decrease.

Intron "Sliding" Is Unlikely—Intron Positions in TPI Are Consistent with Random Insertion. Intron sliding (20–23), which has been invoked to infer homology of introns located at slightly different positions in different organisms (6), is an increasingly unlikely explanation for TPI introns. The exon sizes expected from random intron insertion follow an exponential distribution (8), and the current set of 21 TPI introns is entirely consistent with a random insertion model in which no sliding is necessary. As shown in Fig. 2, the size distribution of 22 "real" exons (the distances between introns in a TPI gene containing all 21 known positions) is similar to that of a randomly generated set of 22 exons. In contrast, for a TPI gene with 12 exons (allowing for three slides; Fig. 3), Gilbert and Glynias (8) suggested that the distribution of distances between introns is "quite different from a random distribution." However, this statement was based on an improper comparison. Gilbert and Glynias (figure 7 of ref. 8) invoked three slides for their "real exons," thereby arbitrarily removing the smallest exons (of 1, 7, and 9 bp) from the real set, yet their random set included this small size class. In fact, for the TPI gene configuration available to Gilbert and Glynias (i.e., with the 14 previously described introns and no slides), exon sizes are compatible with the exponential distribution expected from random insertion ($P > 0.5$, according to a Kolmogorov–Smirnov one-sample test). Similarly, for a gene with all 21 introns now known, exon sizes are also consistent with the exponential distribution ($P > 0.5$).

Although we have allowed for intron sliding in tests of extensity (see below), there is little, if any, empirical evidence supporting such a process in general and none for TPI in particular. Even if sliding were possible, it is difficult to explain the disjunct phylogenetic distribution of several pairs of nearby

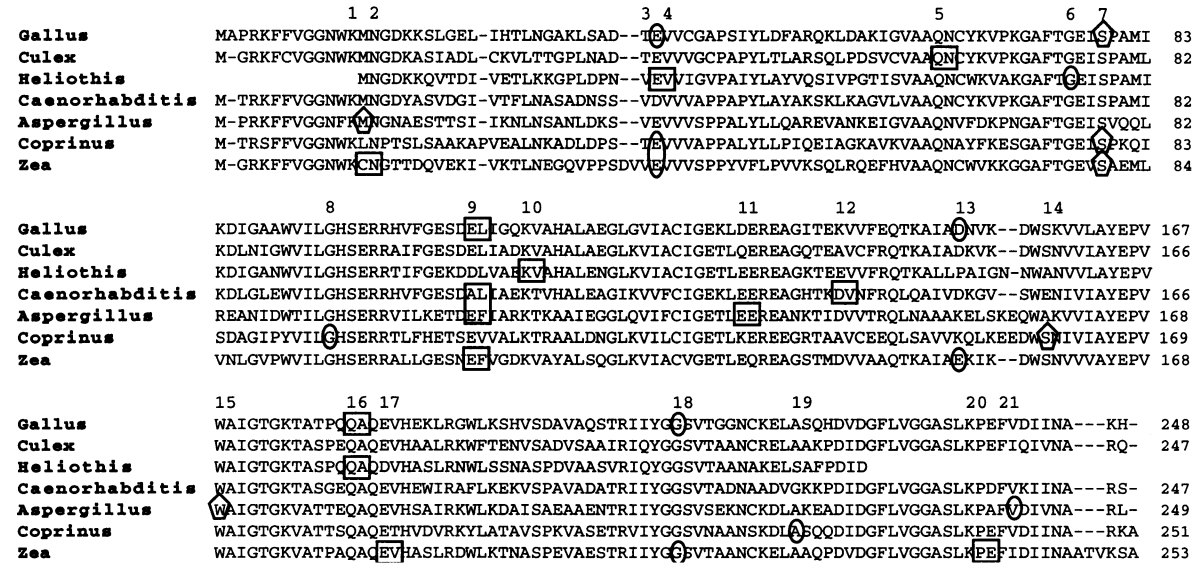

FIG. 1. Alignment of TPI amino acid sequences from genes containing all known intron positions. Intron positions are indicated as boxes where they fall between codons (phase 0) and as ellipses (phase 1) or pentagons (phase 2) where they interrupt codons. Dashes indicate gaps.

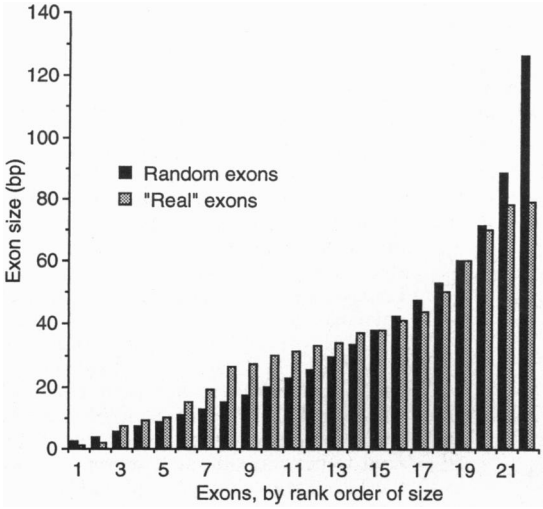

FIG. 2. Comparison of "real" versus randomly generated exon lengths in TPI. The exons are numbered by increasing size. Real exons are the inter-intron distances, in bp, between the 21 known intron positions in TPI, using *Culex* for the reference gene length (741 bp). Random exons are the average inter-intron distances of 100 sets of 21 randomly generated intron positions.

intron positions for which sliding might be invoked. For example, introns at positions 1 and 2 are presumed to be homologous by introns-early advocates (6, 8), but they have a highly disjunct phylogenetic distribution, with many intervening lineages containing neither intron (Fig. 4). Also, four of the six slides considered here (including the three shortest ones) involve a mechanistically improbable change-of-phase with respect to the TPI reading frame (20–23). Moreover, the density of introns in TPI has reached the point at which additional, nonhomologous introns are, and will continue to be, located close to each other by chance alone.

There Is No Correlation Between TPI Exons and Protein Structure. These seven additional introns cast further doubt on the supposed correlations between exons and protein domains (6–8, 13, 14, 17, 22). In addition to eliminating the 12-exon model for the assembly of TPI postulated by introns-early advocates (refs. 6–8 and 15; Fig. 2), these introns further weaken the proposed correspondence between protein structural elements and intron positions in TPI. As shown in Fig. 3, the newly identified introns do not appear to delimit recognized structural features of the TPI protein. Stoltzfus *et al.* (17) recently carried out a statistical analysis of these possible correlations in a TPI gene with 14 distinct intron positions; no significant tendency was found for introns to fall between secondary structural elements (α-helices and β-strands), "Gō modules" (14), or central regions of the single globular domain. Using the methods of Stoltzfus *et al.* (17), we have analyzed the 21-intron TPI "gene" and have found that all correspondences remain statistically insignificant (Table 1). In addition, Stoltzfus *et al.* (17) carried out analyses of extensity (an inverse measure of "compactness") of inferred exon-encoded peptides. For a 12-exon inferred ancestral TPI gene (Fig. 2) they found no more tendency for exons to encode compact peptides than would be expected by chance (Table 2). In contrast, Gilbert and Glynias (8), using a different metric, reported a significant correlation in their separate analysis of extensity of TPI exon-encoded peptides from a 12-exon TPI gene (Table 2). We have carried out analyses of extensity, applying the metrics used by both groups and including the current intron data (Table 2). Allowing for possible intron sliding (6, 20–23) of up to 15 nucleotides, we have considered a number of possible configurations of the inferred ancestral TPI gene from 22 exons (no sliding) to 16 exons (six slides). In all cases, regardless of the metric used, there is no significant tendency for the inferred ancestral exons to encode compact peptides.

Table 1. Correspondence of 21 intron positions to TPI protein structure

	Observed score	Reference score*	*P*
Secondary structure	6.4 bp	6.6 ± 1.4 bp	0.47
Gō modules	13.2 bp	15.4 ± 2.2 bp	0.15
Centrality	15.7 Å	16.2 ± 1.0 Å	0.30

For tests of secondary structure and modules, the scores represent the distance, in base pairs, to the nearest inter-element or inter-module region. For centrality, the scores are the distance, in angstroms, to the center of mass of the TPI protein. The reference scores are the means of scores for 1000 randomly generated sets of 21 intron positions. The *P* value is the probability that a set of 21 randomly generated introns corresponds to the given model of protein structure as well as or better than the set of 21 observed introns. Results presented here use a uniform model for generating reference intron positions. Another model, which generates reference introns using permutation of the inter-intron distances of the observed set, gave similar results for all tests, with none resulting in *P* values <0.05.
*Mean ± SD (randomly placed introns).

Many TPI Introns Were Gained During Recent Eukaryotic Evolution. Parsimony considerations strongly favor an insertional, eukaryotic origin for 16 of the 21 TPI introns. Twelve of 21 introns (marked "recent" in Fig. 4), including all seven introns reported in this study, are found in only a single reported species. The four introns shared between two animal phyla [position 16; ca. 500 million years (myr) old] or between monocots and dicots (positions 2, 17, and 20; ca. 150–200 myr old) are marked as "intermediate." With the single exception of position 16 (which requires but one loss), all 16 recent and intermediate introns can be most parsimoniously explained by a single gain with no subsequent losses; in contrast, an introns-early interpretation requires from 8 to 12 independent losses *at each position.* Parsimony also favors, albeit less strongly, an insertional origin of the five other TPI introns (marked "old" in Fig. 4), which all predate the animal/plant divergence. At least five losses must be invoked at each position to fit the introns-late model (assuming a single gain), whereas an introns-early interpretation would require up to twice as many losses. Further, we predict that a parsimony argument for these old introns being acquired rather than

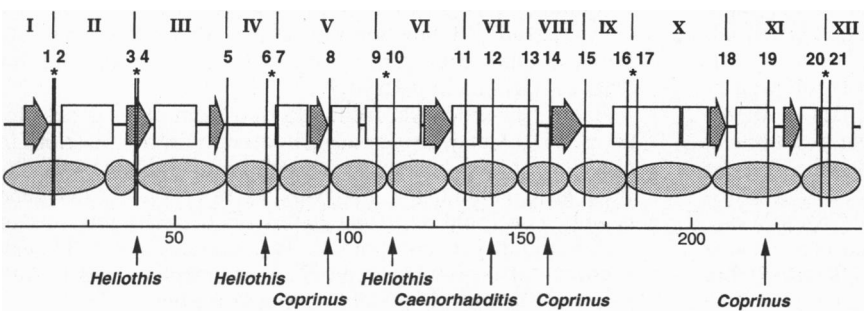

FIG. 3. Comparison of TPI intron positions and protein structural elements. The 21 known intron positions, numbered as in Fig. 1, are represented by long vertical lines. Asterisks mark potential intron "slides." Roman numerals mark the 12 exons of the 11-intron ancestral TPI gene postulated by introns-early advocates (6–8, 14, 15). The reference protein is from *Gallus.* Filled arrows mark β-strands and open boxes indicate α-helices. Gō modules (table 2 of ref. 17) are shown as ellipses. The seven newly identified introns are indicated below the amino acid scale bar.

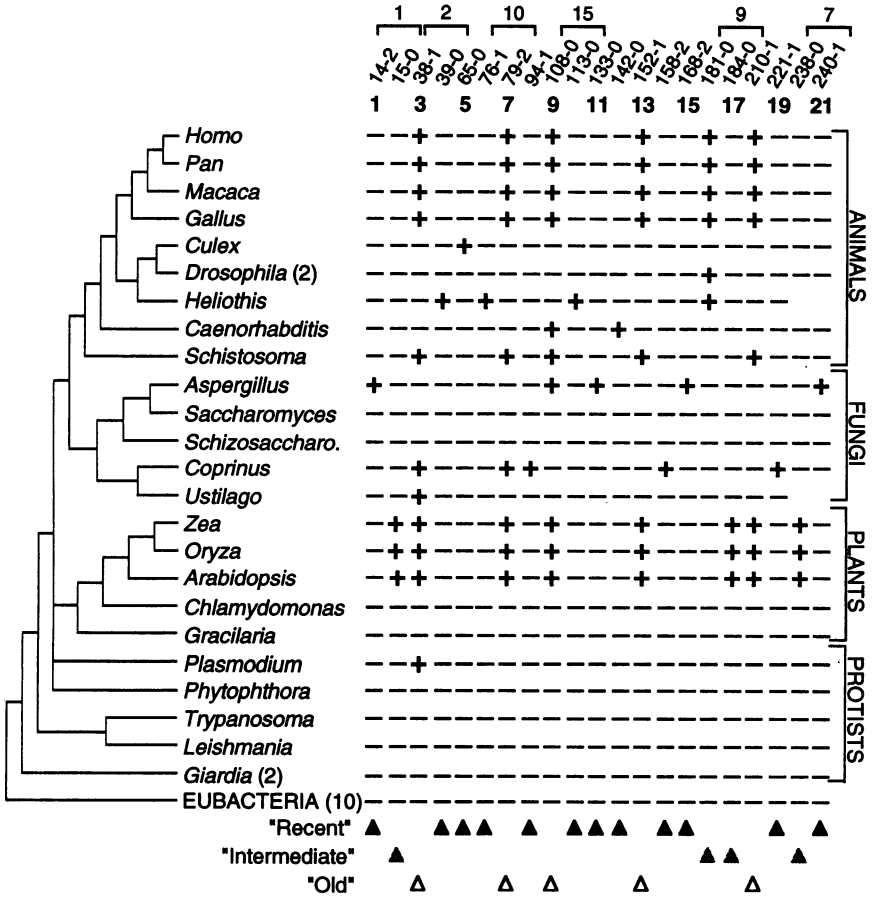

FIG. 4. Phylogenetic distribution of the 21 known TPI introns. Presence of an intron is represented by a plus sign, absence by a minus, and missing data (i.e., partial sequences) by blank space. Intron positions, numbered as in Fig. 1, are annotated with the codon and phase relative to the *Gallus* sequence. Potential intron slides are bracketed, with their distances in base pairs shown above. The cladogram of organismal relationships is based on morphology (especially for animals) and on sequences from rRNA (24, 32, 33), nuclear protein (34), and mitochondrial (B. F. Lang and M. Gray, personal communication) genes. A phylogenetic tree (data not shown) of TPI from the same set of eukaryotes reveals poor resolution of these taxa; hence other data were used to construct the tree shown. Nonetheless, the TPI gene tree is not significantly inconsistent with the tree shown; in particular, there is no compelling evidence to postulate lateral transfer or paralogy of TPI genes. Each taxon represents a single species except where the number of species is given in parentheses. Introns are denoted "old" if present in identical locations in animals and plants, "intermediate" if present in multiple phyla of animals or classes of plants, and "recent" if present in only a single species. All TPI sequences are in GenBank except for *Drosophila simulans* (W. Eanes, personal communication), *Ustilago maydis* (W. Fischer and J.D.P., unpublished data), *Arabidopsis thaliana* (M.-C. Shih, personal communication), *Chlamydomonas reinhardtii* (J.M.L., C.D., and J.D.P., unpublished data), and *Phytophthora infestans* (S. Unkles, J.M.L., and J. Kinghorn, unpublished data).

ancestral will be strengthened as TPI genes are sequenced from more protists, which generally have few, if any, introns (10). Also, TPI is exceptional in the numbers of introns shared between animals and plants (and hence old); two-thirds of the 20+ genes compared between the two groups do not share any introns, while the rest largely have only one or two introns in common (J.M.L. and J.D.P., unpublished data).

Of the 12 recent TPI introns, the intron found in *Culex* (15) is, ironically, now perhaps the best candidate for a relatively recent intron insertion, since, among the several insects examined, it is found only in *Culex* and its sister genus *Aedes* (ref. 25; M.G.T. and V.K.W., unpublished data; W. Fischer and J.D.P., unpublished data). Although cited as confirmation of the exon theory of genes (8, 15), it now seems unlikely that this intron was present in the common ancestor of animals or even insects, much less in the ancestral TPI gene itself. It is difficult to assess the relative recency of the remaining 11 recent introns, since TPI gene sequences from appropriate, closely related organisms have not yet been determined. Many other examples of introns with similarly restricted distributions are known, for example, in genes for GAPDH (19), actin and tubulin (26), and Cu,Zn superoxide dismutase (31).

Further study of these and other presumptively recent intron insertions should help elucidate the process of spliceosomal intron gain, for which three mechanisms have been proposed: (*i*) insertion of a group II intron via reverse self-splicing (21, 22, 27), (*ii*) tandem duplication of exon sequences followed by activation of internal cryptic splice sites (21, 22), and (*iii*) insertion of a transposon that can be spliced out as an intron (28). We are not aware of any evidence supporting the first two mechanisms, but there are numerous cases of latter-day transposon insertions that act as imperfect introns (28), and one particular insertion of the maize Ds element has created a *de novo* intron that is capable of precise excision (29). At 910 bp, the novel intron in *Caenorhabditis* TPI is in the top 5% of *C. elegans* introns in size (K. Steward and T. Blumenthal, personal communication). Although this intron could be a recent insertion that has not experienced substantial diminution, it has no detectable sequence similarity to other *Caenorhabditis* sequences, including known transposons, and its base composition (73% A+T) is typical for *C. elegans* introns (average, 72% A+T; K. Steward and T. Blumenthal, personal communication). The other six TPI introns are similar in base composition and small size to introns found in other genes in *Heliothis* (median intron size, 97 bp; novel introns, 78, 79, and 80 bp) and *Coprinus* (median intron size, 56 bp; novel introns, 63, 71, and 82 bp). These observations, together with a close inspection of the intron sequences and data base searches using them, provide no clues as to their origins.

Conclusion. Although other genes that have been claimed to support the introns-early theory have recently been called into question (17, 19, 23, 30), TPI has remained the exemplar (8, 12). However, our findings undermine TPI as the last major pillar of this theory. We conclude that most TPI introns have been acquired recently in eukaryotic evolution; even the few old introns in TPI are, with but one exception, not present in any examined protist. Our data strongly suggest that as more TPI genes are sequenced, more and more intron positions will be found, most of which will represent recent intron insertions. Thus, instead of being contradicted by TPI, the introns-late theory of eukaryotic gene evolution (9–11) can now cite this gene as evidence in its favor.

Finally, the evidence, from TPI and from other genes (15, 18, 23, 26), of an ever-increasing number of intron positions in a given gene as it becomes sequenced more widely highlights the logical conundrum of trying to infer primordial gene assembly via exon shuffling based on the first few sets of introns found for that gene, as is commonly done. Though discovery of a growing number of recently inserted introns does not rule out the possibility that some limited fraction of

Table 2. Correspondence between inferred TPI exons and extensity

	Observed score	Reference score*	*P*
12 exons (14 known introns, sliding up to 9 bp)			
Diameter†	26.7 Å	26.8 ± 1.4 Å	0.48
Radius†	9.23 Å	9.25 ± 0.43 Å	0.49
Distances‡	Not reported	Not reported	0.039
16 exons (21 introns, sliding up to 15 bp)			
Diameter	22.3 Å	22.5 ± 1.0 Å	0.38
Radius	7.73 Å	7.90 ± 0.30 Å	0.29
Distances	2.56	4.54 ± 2.53	0.24
18 exons (21 introns, sliding up to 9 bp)			
Diameter	20.5 Å	20.5 ± 0.9 Å	0.51
Radius	7.28 Å	7.27 ± 0.25 Å	0.52
Distances	2.28	3.37 ± 2.05	0.34
22 exons (21 introns, no sliding)			
Diameter	17.0 Å	16.9 ± 0.7 Å	0.53
Radius	6.08 Å	6.07 ± 0.20 Å	0.53
Distances	1.68	2.58 ± 1.67	0.35

The "Diameter" metric measures the maximum distance between any two residues in a given exon-encoded-peptide. The "Radius" (of gyration) is the root-mean-square distance of α-carbon atoms from the center of mass of individual exon-encoded-peptides. The "Distances" test, the extensity metric used by Gilbert and Glynias (8) and also repeated by Stoltzfus *et al.* (17), counts the total number of pairwise inter-α-carbon distances >28 Å for each exon-encoded peptide. The "12 exon" case, presented for comparison, excludes the seven introns reported herein and shows the values reported by Stoltzfus *et al.* (ref. 17; †) and Gilbert and Glynias (ref. 8; ‡). The inferred ancestral TPI gene configurations considered here result from differing lengths of intron sliding allowed, from none to 15 bp. The "18 exons" case follows the arbitrarily chosen three-codon rule of Stoltzfus *et al.* (17). For three of the invoked intron slides, the inferred ancestral intron positions are the midpoints of the two nearby intron positions, whereas for the other three slides (positions 3/4, 6/7, and 9/10; Figs. 3 and 4), the presumed ancestral intron position was inferred by parsimony. Extensity metrics require that exons be integral numbers of amino acid codons; therefore, split codons (containing phase 1 and 2 introns) were arbitrarily assigned to the upstream exon (e.g., position 18, 210-1, is coded 211-0) except for the "22 exons" case, in which position 1, 14-2, is coded 14-0 to maintain the distinction from position 2 (15-0). The reference scores are the means of 1000 randomly generated sets of exons. Results presented here are based on a random permutation model of the observed exon sizes for generating the reference set of exons. Another model, which generates reference exons using a lognormal model, gave similar results for all tests, with none resulting in P values <0.05.

*Mean ± SD (randomly placed introns).

extant introns might be ancient, it does make it all but impossible to recognize such introns, let alone prove primordial exon shuffling.

We thank A. Coulson, A. Kerlavage, R. Waterston, and M. Zolan for providing libraries and clones; W. Eanes, W. Fischer, M. Gray, J. Kinghorn, B. F. Lang, M. A. Ragan, M.-C. Shih, S. Unkles, and Y.-H. Zhou for providing unpublished data; T. Inalsingh and M. Ryan for technical assistance; A. Peters for statistical advice; A. Stoltzfus for consultation on protein structure correspondences; and J. Allen, H. Deiderick, C. F. Delwiche, W. Fischer, P. Kuhlman, A. Stoltzfus, C. Tittiger, and M. Zolan for reading the manuscript. J.M.L., C.D., and J.D.-J. were supported by grants to J.D.P. from the National Institutes of Health and the National Science Foundation. M.G.T. and V.K.W. were supported by a grant from the Natural Sciences and Engineering Research Council (Canada).

1. Gilbert, W. (1978) *Nature (London)* **271,** 501.
2. Doolittle, W. F. (1978) *Nature (London)* **272,** 581–582.
3. Darnell, J. E. (1978) *Science* **202,** 1257–1260.
4. Darnell, J. E. & Doolittle, W. F. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 1271–1275.
5. Doolittle, W. F. (1987) *Am. Nat.* **130,** 915–928.
6. Gilbert, W., Marchionni, M. & McKnight, G. (1986) *Cell* **46,** 151–154.
7. Gilbert, W. (1987) *Cold Spring Harbor Symp. Quant. Biol.* **52,** 901–905.
8. Gilbert, W. & Glynias, M. (1993) *Gene* **135,** 137–144.
9. Cavalier-Smith, T. (1985) *Nature (London)* **315,** 283–284.
10. Palmer, J. D. & Logsdon, J. M., Jr. (1991) *Curr. Opin. Genet. Dev.* **1,** 470–477.
11. Cavalier-Smith, T. (1991) *Trends Genet.* **7,** 145–148.
12. Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. & Watson, J. D. (1994) *Molecular Biology of the Cell* (Garland, New York), 3rd Ed.
13. Straus, D. & Gilbert, W. (1985) *Mol. Cell. Biol.* **5,** 3497–3506.
14. Gō, M. & Nosaka, M. (1987) *Cold Spring Harbor Symp. Quant. Biol.* **52,** 915–924.
15. Tittiger, C., Whyard, S. & Walker, V. K. (1993) *Nature (London)* **361,** 470–472.
16. Doolittle, W. F. & Stoltzfus, A. (1993) *Nature (London)* **361,** 403.
17. Stoltzfus, A., Spencer, D. F., Zuker, M., Logsdon, J. M., Jr., & Doolittle, W. F. (1994) *Science* **265,** 202–207.
18. Kersanach, R., Brinkmann, H., Liaud, M.-F., Zhang, D.-X., Martin, W. & Cerff, R. (1994) *Nature (London)* **367,** 387–389.
19. Logsdon, J. M., Jr., & Palmer, J. D. (1994) *Nature (London)* **369,** 526.
20. Patthy, L. (1987) *FEBS Lett.* **214,** 1–7.
21. Rogers, J. H. (1989) *Trends Genet.* **5,** 213–216.
22. Rogers, J. H. (1990) *FEBS Lett.* **268,** 339–343.
23. Stoltzfus, A. & Doolittle, W. F. (1993) *Curr. Biol.* **3,** 215–217.
24. Schlegel, M. (1994) *Trends Ecol. Evol.* **9,** 330–335.
25. Kwiatowski, J., Krawczyk, M., Kornacki, M., Bailey, K. & Ayala, F. J. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 8503–8506.
26. Dibb, N. J. & Newman, A. J. (1989) *EMBO J.* **8,** 2015–2021.
27. Mueller, M. W., Allmaier, M., Eskes, R. & Schweyen, R. (1993) *Nature (London)* **366,** 174–176.
28. Purugganan, M. D. (1993) *Trends Ecol. Evol.* **8,** 239–243.
29. Giroux, M. J., Clancy, M., Baier, J., Ingham, L., McCarty, D. & Hannah, L. C. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 12150–12154.
30. Stoltzfus, A. (1994) *Nature (London)* **369,** 526–527.
31. Kwiatowski, J., Skareeky, D. & Ayala, F. J. (1992) *Mol. Phylogenet. Evol.* **1,** 72–82.
32. Bruns, T. D., Vilgalys, R., Barns, S. M., Gonzalez, D., Hibbett, D. S., Lane, D., Simon, L., Stickel, S., Szaro, T. M., Weisburg, W. G. & Sogin, M. L. (1992) *Mol. Phylogenet. Evol.* **1,** 231–241.
33. Leipe, D. D., Gunderson, J. H., Nerad, T. A. & Sogin, M. L. (1993) *Mol. Biochem. Parasitol.* **59,** 41–48.
34. Baldauf, S. L. & Palmer, J. D. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 11558–11562.