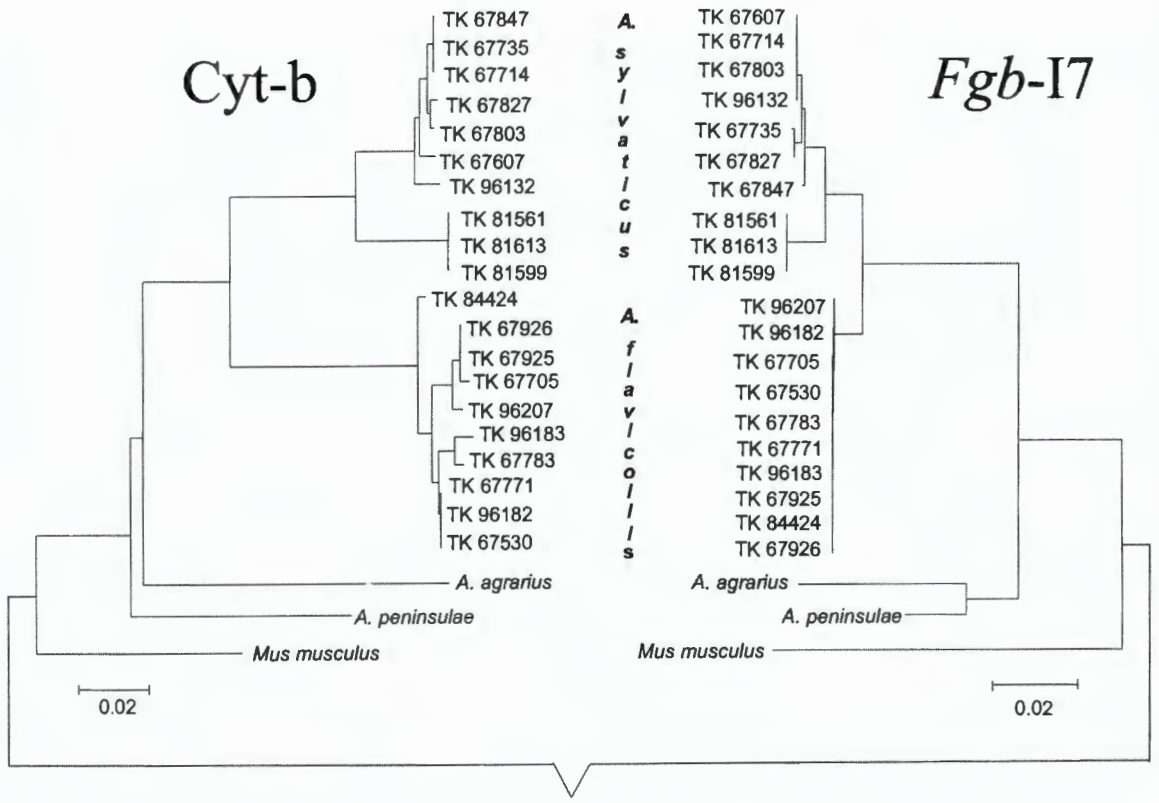
OCCASIONAL PAPERS

Museum of Texas Tech University

Number 219 12 May 2003



INTRON 7 (*FGB*-I7) OF THE FIBRINOGEN, B BETA POLYPEPTIDE (*FGB*): A NUCLEAR DNA PHYLOGENETIC MARKER FOR MAMMALS

Editor's Comment: Data from the mitochondrial genome continue to accumulate that indicate major subdivisions between and among populations currently assigned to single species. Understanding the importance of these subdivisions is significant because if more than a single biological species is present then the biodiversity of these so called "species" has been underestimated. Mitochondrial sequence data are easily obtained and it is our opinion that they have a reasonable probability level of accurately estimating species boundaries for a substantial number of undetected species. As we all know, however, Biology is not an exact science and there will be exceptions to all rules and no simple equation will accurately predict the species level significance of distance values from a given mitochondrial gene. Clearly there will be numerous examples where unrecognized species exist in mammals and we are left with the dilemma of how to tease from the genetic data, the level of resolution that will permit us to recognize distinct biological species as opposed to polymorhpisms within a single biological species. Data from a non-mitochondrial source are needed to better understand how predictable the mtDNA sequences are in documenting species subdivisions. Information from several sources can be used to address this issue. First, detailed field studies on gene flow would be ideal, but such studies are labor intensive and in many cases not possible. Second, classical morphology could be employed, but in these cases the morphology is not too obvious or the populations with the divergent mtDNA sequences would not be considered conspecific. Third, sequence data from a nuclear fragment can be employed to provide an independent estimate of divergence. Most nuclear DNA fragments evolve sufficiently slowly that they do not provide adequate divergence for resolution between closely related species. In this paper, Wickliffe, et al. explores the utility of intron 7 of the beta fibrinogen locus for application where the data from the mitochondrial genome suggest a species level divergence may be present. Within the limited sample available from this intron there is evidence that variation within closely related species may be adequate to contribute to better understanding these species/subspecies level problems. Finally, if the data from the nuclear and mitochondrial genome are congruent, then the conclusions become highly probable because the data sets from the two genomes are not linked, thereby avoiding most types of potential bias.

RJB

Front cover: Opposing neighbor-joining tree for *Apodemus*, generated based on the Tamura Nei model. A - *Mtcyb* (392 bp) tree. B - *Fgb*-I7 tree. *Mus musculus* is the outgroup specimen in both A and B. For the scaling purpose few more individuals of *Apodemus* species and *Mus musculus* were added.

INTRON 7 (*FGB*-I7) OF THE FIBRINOGEN, B BETA POLYPEPTIDE (*FGB*): A NUCLEAR DNA PHYLOGENETIC MARKER FOR MAMMALS

JEFFREY K. WICKLIFFE, FEDERICO G. HOFFMANN, DARIN S. CARROLL, YELENA V. DUNINA-BARKOVSKAYA, ROBERT D. BRADLEY, AND ROBERT J. BAKER

Systematic and molecular evolutionary studies have benefited from the use of DNA sequence markers for phylogenetic reconstruction. Most molecular phylogenies have used maternally inherited mitochondrial (mt) or chloroplast (cp) DNA sequences. The sequence organization, molecular and functional characteristics of these relatively small, non-recombinant genomes has provided a framework for understanding the evolution of DNA sequences and genetic relationships among taxa. However, matrilineal phylogenies provide evolutionary estimates that reflect only uniparental patterns and gene tree relationships. With advances in biotechnology and genomic elucidation, the discovery and development of nuclear DNA sequence markers for phylogenetic reconstruction is now possible for practicing systematists. Nuclear DNA sequences provide genetic histories independent of mtDNA or cpDNA phylogenies and a capitulation of diparental histories when considering sexually reproducing organisms. Nuclear DNA sequences applied thus far to phylogenetic questions have primarily been from introns. This is because introns are believed to primarily evolve in a neutral manner and because exon sequences generally evolve much more slowly than introns (Baker et al. 2000). However, most of the introns that have been investigated to date are relatively invariant at or below the level of species rendering them of limited utility for comparison to mtDNA phylogenies targeting these taxonomic levels. A notable exception to both of these widely held views is illustrated by DeWoody (1999). In addition, introns often evolve by both nucleotide and insertion/deletion (indel) polymorphisms complicating the process of applying models of DNA sequence evolution. Therefore, continued development and comparative analyses seek those nuclear loci which are compatible with rapidly evolving mtDNA genes and currently applied models of DNA sequence evolution. Recently, intron 7 (*Fgb*-I7) of the fibrinogen, B beta polypeptide gene (*Fgb*, single-copy) was examined in two orders of birds and a generic complex of southeast Asian pit vipers (Prychitko and Moore 1997, Prychitko and Moore 2000, Giannasi et al. 2001, Johnson et al. 2001). These studies indicate *Fgb*-I7 appears to evolve in a neutral fashion primarily through nucleotide substitutions and at a sufficient rate to complement corresponding mtDNA species phylogenies.

We developed and applied PCR and sequencing primers for *Fgb*-I7 (*Fgb* maps to chromosome 3, cM position 48.2 in *Mus musculus*) in 2 orders of mammals (Chiroptera and Rodentia) to examine amplification universality and apparent phylogenetic concordance with existing mtDNA gene (cytochrome *b*-*Mtcyb*) phylogenies for congeneric species. The following presentation of the technical aspects of the development and application of this nuclear intron in these taxa is principally designed to centralize the oligo sequences, thermal profiles, and reagent properties of the specific PCRs in one source. Therefore, the phylogenetic conclusions resulting from the associated studies will not be discussed herein.

MATERIALS AND METHODS

Primers were designed from a multiple sequence alignment of the conserved coding region from several mammal species (*Mus musculus*, *Rattus norvegicus*, *Homo sapiens*, *Bos taurus*) and from preliminary sequence data obtained from the taxa examined using the exon-anchored primers of Prychitko and Moore (1997). The Vector NTT Suite 6.0 software was used to analyze all oligos (InforMax Inc., Bethesda, MD). All cycle sequencing reactions, regardless of the primer used, were performed using Big DyeJ versions 2.0 and 3.0 chemistries according to the manufacturer's recommendations (Applied Biosystems, Foster City CA). DNA sequence chromatograms were proofed in Sequencher ver. 3.1 or VectorNTI Suite ver. 6.0. Multiple sequence alignments were generated in VectorNTI Suite ver. 6.0.

Two chiropteran genera, *Glossophaga* (n = 8 species, 17 total taxa) and *Carollia* (n = 5 species, 13 total taxa) and two rodent genera, *Sigmodon* (n = 10 species, 21 total taxa) and *Apodemus* (n = 3 species, 18 total taxa) were analyzed. Each of these analyses also included from 1 to 3 additional genera used to root phylogenies and assess primer utility and *Fgb* intron sequence characteristics. Primer sequences, thermal-cycling profiles, and reagent constitutions used in the PCR for each genus are presented in Table 1.

Nucleotide composition, average substitution profiles, and the phylogenetic consistency index (CI) were calculated for *Mtcyb* and *Fgb* datasets. *Mtcyb* and *Fgb* pairwise, uncorrected *p* distances were statistically compared using correlation analysis

Table 1. Sequences for PCR and cycle-sequencing primers and PCR conditions (i.e. thermal profile, hardware, and reagents) used to generate Fgb DNA sequences for mammalian genera. Primers superscripted by an "a" were used for the PCR and primers superscripted by a "b" were used for sequencing reactions. All PCR were performed in 50 ul volumes. D = denaturation, A = annealing, E = extension.

Genus	PCR[a] and Cycle-Sequencing Primers[b]	Thermal Profile and Cycler	PCR Reagents
(≤300 ng genomic DNA)	(Oriented 5-prime to 3-prime)	(Perkin-Elmer 480)	
Glossophaga	FIB-BI7U*[a,b]: GGAGAAAACAGGACAATGACAATTCAC Fgb-17L-Rod[a,b]: ATGTCCCAGCTGTAAAGGCCACCC	35 cycles of D-93°C-30seconds, A-56°C-30s, E-72°C-140	0.2mM-dNTPs 1.5mM-$MgCl_2$ 5.0µl-10xbuffer 1.5U-enzyme 0.54µM-primer
Carollia	FIB-BI7U[a,b] Fgb-I7L-Rod[a,b]	35 cycles of D-94°C-40s, A-53°C-45s, E-72°C-90s	0.2mM-dNTPs 2.5mM-$MgCl_2$ 5.0µ1-10x buffer 1.5U-enzyme 0.54µM-primer
Sigmodon	Fgb-I7U-Rattus[a,b] GGGGAGAACAGAACCATGACCATCCAC 300F[b]:CAGCAACCAGAGGACATCTCCCTG Fgb-I7L-Rattus: ACCCCAGTAFTATCTGCCATTCGGATT	35 cycles of D-94°C-40s, A-53°C-45s, E-72°C-90s	0.175mM-dNTPs 1.25-$MgCl_2$ 4.8ml-10x buffer 2.5U-enzyme .054Um-primer
Apodemus	Fgb-I7U-Rattus[a,b]- Fgb-I7L-Rattus[a,b]- Apo-intU[b]:AGACAGCYACCCAAAGAT Apo-intL[b]:ATCTTTGGGTRGCTGTCT	35 cycles of D-94°C-40s, A-53°C-45s, E-72°C-90s	0.2mM-dNTPs 0.2mM-$MgCl_2$ 5.0µl-10x buffer 1.5U-enzyme 0.54µM-primer

*reported in Prychitko and Moore (1997)

(Pearson's r^2) in the SPSS ver 11.0 for Windows program. An á < 0.01 was used for determining significance. The slope (m) of a least squares linear fit was estimated using the SPSS software. Phylograms for *Fgb* and *Mtcyb* datasets were compared using partition metrics generated in the COMPONENT version 2.0 program (Page 1993). The partition metric (PM) is defined as the number of clusters found in 1 tree or the other but not both (Day 1985, Penny and Hendy 1985). Therefore, a tree compared to itself will have a PM = 0. Partition metrics from 1000 random trees generated from an equivalent number of leaves (i.e. terminal taxa) for each respective dataset were also calculated. This random distribution of PMs can be used to statistically assess the structural similarities between the two trees. If the PM from the *Mtcyb* and *Fgb* comparison is below the smallest PM observed at a frequency of 5% or more from the random trees comparisons, we can assume there is significant similarity ($p < 0.05$ of the observed similarity being random) between the *Mtcyb* and *Fgb* phylograms.

*Table 2. Nucleotide composition, nucleotide substitution profiles (TS = transition, TV = transversion), the phylogenetic consitency index (CI), partition metrics (PM), correlation coefficients (Pearson's r^2) and slope estimates (*m*) for uncorrected* p *distances are provided for both* Mtcyb *and* Fgb *analyses among mammalian genera.*

	%A	*%C*	*%G*	*%T*	*TS*	*TV*	*CI*	*PM*	r^2	*m*
		Glossophaga						12 (24*)	0.74[a]	0.51
Fgb	28.7	20.5	19.9	31.0	12	5	0.91			
Mtcyb	28.7	26.1	13.8	31.4	111	29	0.54			
		Carollia						12 (16)	0.79[a]	0.59
Fgb	31.6	20.7	18.6	29.1	10	3	0.94			
Mtcyb	28.0	30.4	13.9	27.7	93	22	0.74			
		Sigmodon						22 (32)	0.78[a]	0.37
Fgb	33.0	21.8	16.8	28.3	8	5	0.95			
Mtcyb	27.3	28.6	12.9	31.3	102	33	0.53			
		Apodemus						26 (26)	0.70[a]	0.58
Fgb	27.1	24.7	21.0	27.1	14	8	0.97			
Mtcyb	31.4	26.4	12.3	29.9	66	28	0.85			

*values in parentheses represent the lower bound on the distribution of partition metric distances (frequency <0.05) generated from 1000 random trees.
[a]statistically significant at α<0.01

RESULTS

The *Fgb*-I7 DNA sequences obtained for all specimens in each genus contained both the highly conserved flanking sequences characteristic of the *Fgb* exons upstream and downstream and the splice junction sequences (i.e. GTnYnAG) found in all vertebrates thus far studied. This suggests that we successfully amplified the *Fgb*-I7 locus. The *Fgb*-I7 varied in length (511 to 628 base pairs [bp]) among all taxa included in this study and was consistently shorter than the DNA sequence obtained from birds (> 900 base pairs [bp]) and vipers (927 bp). While a few indels were observed, nucleotide substitutions constituted the predominant type of polymorphism within each genus analyzed. A summary of basic DNA sequence characteristics for each genus is provided in Table 2. A notable A:T bias was observed in all genera consistent with the nucleotide compositional bias reported for birds and vipers (Prychitko and Moore 2000, Giannasi et al. 2001). Consistency indices for the *Fgb*-I7 datasets were substantially higher than those for the *Mtcyb* datasets. Correlation analyses indicated that genetic distances from both datasets were significantly ($p < 0.01$) correlated for each genus investigated. Pearson's r^2 ranged from 0.70 to 0.78. Estimates of linear slope, m, ranged from 0.37 to 0.59. Partition metrics for each comparison of the 4 mammalian genera were below the 5% frequency level for PMs generated from random trees.

DISCUSSION

These preliminary analyses indicate *Fgb*-I7 is potentially useful for identifying species of mammals and for complementing existing phylogenies. The evolutionary rate of *Fgb*-I7 appears to be slower than the rate estimated for *Mtcyb* as indicated by the lower estimates of genetic distance for *Fgb*-I7 and corresponding values of m (< 1). Consistency indices indicate little homoplasy within the *Fgb* datasets in contrast to the *Mtcyb* datasets which are slightly compromised because of probable saturation effects. Genetic distance estimates are also significantly correlated and a substantial amount of the variation observed in each set of distance estimates is explained by these correlations. These correlated relationships are best explained by a common evolutionary history because otherwise, variation present within each dataset should be randomized with respect to the other (largely due to unlinked, neutrally evolving DNA sequences). Partition metrics for *Fgb*-I7 and *Mtcyb* phylograms are consistently lower than the PMs generated from random tree comparisons. This suggests that the nuclear and mitochondrial DNA phylograms are not significantly different. Finally, polymorphisms primarily result from nucleotide substitutions and not indels. This allows for the reconstruction of straightforward sequence alignments and the subsequent application of standard models of DNA sequence evolution to be applied to these datasets for phylogenetic estimation. All of these properties support the utility of *Fgb*-I7 as a phylogenetic marker and as an independent nuclear locus for examining congruence with mitochondrial gene trees.

Further evaluation of *Fgb*-I7 is needed to provide a more complete database of comparative material before its full usefulness can be appreciated. As generating DNA sequence data becomes more efficient, *Fgb*-I7 will likely be used as one of several independent and dependent markers used for phylogeny reconstruction. This is because it simply is not possible for a single marker to reveal accurate phylogeny in all possible cases. However, while the noncoding property of intron sequences is likely responsible for their increased evolutionary rate and attractiveness as near-neutral DNA markers, it is a limitation when researchers are interested in protein evolution, protein phylogenies, or examining natural selection at the molecular/biochemical level. We make this caveat because a considerable amount of current and emerging research is addressing these types of evolutionary questions in a phylogenetic context. Whether or not *Fgb*-I7 is useful for species' level phylogenetic reconstruction in a wide variety of applications remains to be seen. However, it now has been demonstrated to be useful in studies of 3 vertebrate classes. In this study, we give the primer sets and conditions which provide molecular systematists and evolutionary biologists an initial point for applying a new tool, *Fgb*-I7, with which to investigate the evolutionary relationships among species of mammals.

LITERATURE CITED

Baker, R. J., C. A. Porter, J. C. Patton, and R. A. Van Den Bussche. 2000. Systematics of bats of the family Phyllostomidae based on *Rag*2 DNA sequences. Occasional Papers, Museum of Texas Tech University 202: 1-16.

Day, W. H. E. 1985. Optimal algorithms for comparing trees with labeled leaves. Journal of Classification 1: 7-28.

DeWoody, J. A. 1999. Nucleotide variation in the p53 tumor-suppressor gene of voles from Chernobyl, Ukraine. Mutation Research/Genetic Toxicology and Environmental Mutagenesis 439:25-36.

Giannasi, N., A. Malhotra, and R. S. Thorpe. 2001. Nuclear and mtDNA phylogenies of the *Trimeresurus* complex: implications for the gene versus species tree debate. Molecular Phylogenetics and Evolution 19: 57-66.

Johnson, K. P., S. de Kort, K. Dinwoodey, A. C. Mateman, C. ten Cate, C. M. Lessells, and D. H. Clayton. 2001. A molecular phylogeny of the dove genera *Streptopelia* and *Columba*. Auk 118: 874-887.

Page, R. D. M. 1993. User's manual for COMPONENT, Version 2.0. The Natural History Museum, London.

Penny, D., and M. D. Hendy. 1985. The use of tree comparison metrics. Systematic Zoology 34: 75-82.

Prychitko, T. M., and W. S. Moore. 1997. The utility of DNA sequences of an intron from the â-fibrinogen gene in phylogenetic analysis of woodpeckers (Aves: Picidae). Molecular Phylogenetics and Evolution 8: 193-204.

Prychitko, T. M., and W. S. Moore. 2000. Comparative evolution of the mitochondrial cytochrome *b* gene and nuclear â-fibrinogen intron 7 in woodpeckers. Molecular Biology and Evolution 17: 1101-1111.

Addresses of authors:

JEFFREY K. WICKLIFFE

Department of Biological Sciences
Texas Tech University
Lubbock, TX 79409-3131
jkwickli@utmb.edu

FEDERICO G. HOFFMANN

Department of Biological Sciences
Texas Tech University
Lubbock, TX 79409-3131
fhoffman@ttu.edu

DARIN S. CARROLL

Department of Biological Sciences
Texas Tech University
Lubbock, TX 79409-3131
dcarroll@cdc.gov

YELENA V. DUNINA-BARKOVSKAYA

Department of Biological Sciences
Texas Tech University
Lubbock, TX 79409-3131
ydunina@ttu.edu

ROBERT D. BRADLEY

Department of Biological Sciences
Museum of Texas Tech University
Texas Tech University
Lubbock, TX 79409-3131
robert.bradley@ttu.edu

ROBERT J. BAKER

Department of Biological Sciences
Museum of Texas Tech University
Texas Tech University
Lubbock, TX 79409-3131
robert.baker@ttu.edu

Publications of the Museum of Texas Tech University

Institutional subscriptions are available through the Museum of Texas Tech University, attn. NSRL Publications Secretary, Box 43191, Lubbock, TX 79409-3191. Individuals may also purchase separate numbers of the Occasional Papers directly from the Museum of Texas Tech University.

Layout and Design: Jacqueline B. Chavez
Cover Design: Robert J. Baker

This book was set in Times New Roman and printed on acid-free paper that meets the guidelines for permanence and durability of the Committee on Production Guidelines for Book Longevity of the Council on Library Resources.

Printed: 12 May, 2003

Library of Congress Cataloging-in-Publication Data

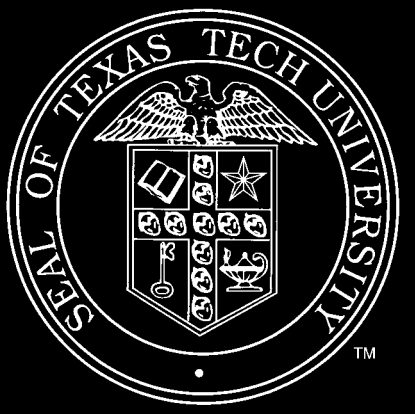
Occasional Papers, Number 219
Series Editor: Robert J. Baker

INTRON 7 (*FGB*-I7) OF THE FIBRINOGEN, B BETA POLYPEPTIDE (*FGB*): A NUCLEAR DNA PHYLOGENETIC MARKER FOR MAMMALS

By: Jeffrey K. Wickliffe, Federico G. Hoffmann, Darin S. Carroll, Yelena V. Dunina-Barkovskaya, Robert D. Bradley, and Robert J. Baker

ISSN 0149-175X

Museum of Texas Tech University
Lubbock, TX 79409-3191 USA
(806)742-2442


SEAL OF TEXAS TECH UNIVERSITY
TM