Retroviruses drive the rapid evolution of mammalian APOBEC3 genes

Jumpei Ito*, Robert J. Gifford*, and Kei Sato*·c,1

*Division of Virology, Department of Infectious Disease Control, International Research Center for Infectious Diseases, Institute of Medical Science, The University of Tokyo, Tokyo 1088639, Japan; †Medical Research Council-University of Glasgow Centre for Virus Research, University of Glasgow, Glasgow G61 1QH, United Kingdom; and ‡Core Research for Evolutionary Medical Science and Technology (CREST), Japan Science and Technology Agency, Saitama 3220012, Japan

Edited by John M. Coffin, Tufts University, Boston, MA, and approved November 12, 2019 (received for review August 15, 2019)

APOBEC3 (A3) genes are members of the AID/APOBEC gene family that are found exclusively in mammals. A3 genes encode antiviral proteins that restrict the replication of retroviruses by inducing G-to-A mutations in their genomes and have undergone extensive amplification and diversification during mammalian evolution. Endogenous retroviruses (ERVs) are sequences derived from ancient retroviruses that are widespread mammalian genomes. In this study we characterize the A3 repertoire and use the ERV fossil record to explore the long-term history of coevolutionary interaction between A3s and retroviruses. We examine the genomes of 160 mammalian species and identify 1,420 AID/APOBEC-related genes, including representatives of previously uncharacterized lineages. We show that A3 genes have been amplified in mammals and that amplification is positively correlated with the extent of germline colonization by ERVs. Moreover, we demonstrate that the signatures of A3-mediated mutation can be detected in ERVs found throughout mammalian genomes and show that in mammalian species with expanded A3 repertoires, ERVs are significantly enriched for G-to-A mutations. Finally, we show that A3 amplification occurred concurrently with prominent ERV invasions in primates. Our findings establish that conflict with retroviruses is a major driving force for the rapid evolution of mammalian A3 genes.

mammal | APOBEC | gene amplification | endogenous retrovirus | evolutionary arms race

Activating-induced cytidine deaminase/apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (AID/APOBEC) superfamilies are cellular cytosine deaminases that catalyze cytosine-to-uracil (C-to-U) mutations. AID/APOBEC family proteins contain a conserved zinc-dependent catalytic domain (Z domain) with the HxEC/PxExC motif and are closely associated with important phenomena found in vertebrates such as immunity, malignancy, metabolism, and infectious diseases (reviewed in refs. 1 and 2). For instance, AID induces somatic hypermutation in B cells and promotes antibody diversification (2), and APOBEC1 (A1) regulates lipid metabolism by enzymatically editing the mRNA of apolipoprotein B gene (3). The physiological roles of APOBEC2 (A2) and APOBEC4 (A4) remain unknown, but APOBEC3 (A3) genes are known to encode antiviral factors that restrict the replication of retroviruses (4) and other viruses (5–7).

While most AID/APOBEC family genes are conserved in vertebrates, A3 genes are specific to placental mammals (1). Furthermore, whereas AID, A1, A2, and A4 genes are singly encoded in each vertebrate including mammals, dramatic expansion of the A3 repertoire occurred in many mammalian lineages, including primates (8). A3 genes are grouped into 3 classes (A3Z1, A3Z2, and A3Z3) on the basis of their conserved Z domain sequences (4, 8, 9). For example, human A3 genes are composed of 7 paralogs (A3A, A3B, A3C, A3D, A3F, and A3H). Of these, A3A, A3C, and A3H (which in other mammals are referred to as A3Z1, A3Z2, and A3Z3, respectively) contain a single Z domain, while the other 4 genes harbor double Z domains: A3Z2, A3Z1 for A3B and A3G and A3Z2, A3Z3 for A3D and A3F (8, 9).

The conflict between human A3G protein and HIV type 1 (HIV-1) has been studied particularly intensively. Human A3G proteins are incorporated into HIV-1 particles and enzymatically induce C-to-U mutations in viral cDNA, causing guanine-to-adenine (G-to-A) mutations in the viral genome (10, 11). A3G-mediated mutations lead to the accumulation of lethal mutations and ultimately abolish viral replication. On the other hand, an HIV-1–encoding protein, viral infectivity factor (Vif), counteracts this antiviral action by degrading A3G in a ubiquitin-proteasome–dependent manner (4). Such conflicts between A3 proteins and modern viruses (particularly retroviruses) have been reported in a broad range of mammalian species and viruses infecting them (reviewed in ref. 9), and consistent with this, A3 genes contain strong signatures of diversifying selection (12–14).

Endogenous retroviruses (ERVs) are retrotransposon lineages that are thought to have originated from ancient exogenous retroviruses via infection of germline cells (15, 16). ERVs occupy a substantial fraction of mammalian genomes, demonstrating extensive germline invasion by retroviruses. To combat ERVs and other intragenomic parasites, mammals have developed defense systems such as Kruppel-associated box domain-containing (KRAB) zinc finger proteins (17) and PIWI-interacting RNAs (18). A3 proteins have been shown to suppress the replication of reconstructed ERVs in cell cultures (15, 19) and in a transgenic mouse model (20). Furthermore, previous studies identified the signature of A3-mediated G-to-A mutations in ERVs indicating that ancient retroviruses experience attacks by A3 proteins (15, 16, 19, 21). In this study, we examine the history of evolutionary

**Significance**

It is thought that evolution of antiviral genes has been shaped over the long term by antagonistic interactions with viruses, but in most cases this is challenging to investigate. In this study we examine the evolution of A3 genes—antiviral genes that target retroviruses by inducing mutations in their genomes. We demonstrate that ancient, fossilized retrovirus sequences in mammalian genomes contain clear signatures of A3-mediated mutation and provide several additional lines of evidence that A3 evolution has been driven by long-running conflicts with ancient retroviruses.

Author contributions: J.J., R.J.G., and K.S. designed research; J.J. and R.J.G. performed research; R.J.G. contributed new reagents/analytic tools; J.I. analyzed data; and J.I., R.J.G., and K.S. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

Data deposition: The data, associated protocols, code, and materials in this study are available at https://giffordlabcvr.github.io/A3-Evolution/.

*To whom correspondence may be addressed. Email: ksato@ims.u-tokyo.ac.jp.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1914183116/-/DSSupplemental.

https://www.pnas.org/cgi/doi/10.1073/pnas.1914183116 PNAS Latest Articles | 1 of 9
interaction between ERVs and A3 genes via genomic analysis of 160 mammalian species.

Results

Identification and Classification of Mammalian AID/APOBEC Family Genes. We screened whole genome sequence (WGS) data of 160 mammalian species in silico and extracted 1,420 sequences disclosing homology to the conserved Z domains of AID/APOBEC family genes (SI Appendix, Fig. S1 and Datasets S1–S3). Phylogenetic reconstructions revealed that these Z domain loci group into 9 clades, 7 of which represent the canonical AID/APOBEC lineages (AID, A1, A2, A3Z1, A3Z2, A3Z3, and A4) (Fig. 1 A and B). We also identified additional, previously uncharacterized lineages, designated UA1 and UA2 (Fig. 1 A and B). UA1 genes were only found in basal eutherian mammal groups: Afrotherians (elephants, tenrecs, and sea cows) and xenarthrans (armadillos). UA2 genes were only found in marsupials (infraclaus Marsupialia) (Fig. 1C). These phylogenetic relationships were supported by multiple methods (Fig. L4 and SI Appendix, Fig. S2A). In addition, HxE and PCxCxC motifs corresponding to the canonical catalytic domain of AID/APOBEC proteins were found in UA1 and UA2 gene sequences (SI Appendix, Fig. S2B). The UA1 and UA2 genes contain signatures of purifying selection (SI Appendix, Fig. S2C) indicating they are protein-coding members of the AID/APOBEC family. Indeed, the UA2 gene in opossum (Monodelphis domestica) was annotated as AP024436 in a previous study (22).

As summarized in Fig. 1B, we detected 157 AID, 166 A1, 157 A2, 266 A3Z1, 362 A3Z2, 146 A3Z3, 153 A4, 9 UA1, and 4 UA2 genes in 160 species of mammalian genomes. Interestingly, A3Z1 and A3Z2 genes were highly amplified, while the other family genes were not (Fig. 1 B and C). We also found that some sequences, particularly those of A3 genes, were pseudogenized (Fig. 1B). The numbers of A3 Z domains were different among species. In particular, A3Z1 and A3Z2 genes in Perissodactyla, Chiroptera, Primates, and Afrotheria were highly amplified (Fig. 1C and SI Appendix, Fig. S3). Consistent with previous reports (12, 23, 24), canonical A3 genes were not detected in marsupials or monotremes (order Monotremata). Furthermore, A3Z1 was commonly absent in Rodentia, while A3Z2 was absent in Strepsirhini and Microchiroptera. Amplification of A3Z3 genes was not detected in any mammalian groups except for Carnivora (carnivores), in which duplicated A3Z3 genes were almost entirely pseudogenized (SI Appendix, Fig. S4).

Evolution of Mammalian A3 Genes Under Strong Selection Pressures. We used comparative genomic approaches to investigate the evolutionary history of mammalian A3 genes. As shown in Fig. 2A, the positional conservation (Shannon entropy) scores in A3Z1, A3Z2, and A3Z3 genes tended to be much higher than those found in other AID/APOBEC family genes, indicating strong diversifying selection. We detected codon sites evolving under diversifying selection by calculating dN/dS ratios using the branch-site model (25). Although the catalytic domains, which are composed of HxE and PCxCxC motifs (1, 2, 4), were highly conserved among the 7 AID/APOBEC family proteins, we detected the signature of diversifying selection at numerous sites (Fig. 2B). Comparisons to human A3A (A3Z1 ortholog in primates) (26), A3C (A3Z2 ortholog in primates) (27), and A3H (A3Z3 ortholog in primates) (28) revealed that these sites are preferentially detected in a structural region called loop 7, which recognizes substrate nucleic acids (Fig. 2B). Furthermore, most of the sites under diversifying selection are located on the protein surface (Fig. 2B).

Investigation of amplified A3 loci revealed that the majority of A3 genes are encoded in the canonical A3 genomic locus (8, 9), flanked by the CBX6 and CBX7 genes (Fig. 3A and Dataset S4), indicating that amplification of A3 genes has mainly occurred via tandem gene duplication. However, there are exceptions to this rule: 3 primate species, Saimiri boliviensis, Aotus nancymaeae, and Otomeur garnettii, were found to encode more A3 loci outside the canonical locus than within it (Fig. 3B). The A3 genes in these 3 primates were mostly encoded at entirely distinct loci (Fig. 3C) and exhibit double-domain (A3Z2–A3Z1) and intronless structures (SI Appendix, Fig. S5A and Dataset S5) indicating they likely originated via retrotransposition of spliced mRNA (29). These retrotransposed A3 genes in New World monkeys were more closely related to the human A3G gene than the other double-domain A3 genes in humans (SI Appendix, Fig. S5B). Although most were pseudogenized (Fig. 3D), some retain relatively long ORFs (SI Appendix, Fig. S5C). In particular, 1 of the retrotransposed A3 genes in A. nancymaeae (referred to as “outside #3”) retains a full-length ORF (SI Appendix, Fig. S5C). Indeed, this gene is annotated in the Ensembl gene database (http://www.ensembl.org; Release 97; ENSANAG00000031271). Moreover, analysis of public RNA-sequencing (RNA-Seq) data revealed that mRNA of outside #3 is expressed in a broad range of tissues in A. nancymaeae (SI Appendix, Fig. S5D). Taken together, these data show that A3G-like genes have been amplified via retrotransposition in New World monkeys, and some of these amplified genes are likely functional.

ERVs Evidence a Long-Running Conflict Between Retroviruses and A3 Genes. To explore the impact of A3 activity on ERVs and their ancient exogenous ancestors, we performed comparative analysis of transposable elements (TEs) in 160 mammalian genomes. As shown in Fig. 4A and SI Appendix, Fig. 4B, the TE composition of mammalian species varies with respect to the proportions of DNA transposons, SINEs, LINEs, and ERVs. To investigate the accumulation level of G-to-A mutations in ERVs, we measured the strand bias of the G-to-A mutation rate in ERVs and other TEs. Since A3 proteins selectively induce G-to-A mutations on the strand bias of the G-to-A mutation rate in ERVs and other TEs. Since A3 proteins selectively induce G-to-A mutations, we considered the A3 genes as a model for such an effect. As shown in Fig. 4C, ERVs in the human genome preferentially exhibited GG-to-GA or GA-to-AA mutations, consistent with the reported preferences of human ASG (GG-to-AG) and A3D, A3F, and ASH (GA-to-AA mutations) (30, 33–39). Additionally, some ERVs exhibited G-to-A hypermutation (Fig. 4D).

To explore the potential impact of A3 gene amplification on ERVs, we first assessed the accumulation level of G-to-A mutations across all mammalian ERVs (SI Appendix, Fig. S7), then examined the association between 1) accumulation of G-to-A mutations in ERVs and 2) the number of A3 Z domains. This revealed a strong positive correlation (Fig. 4E) (Pearson’s correlation coefficient = 0.69, P < 1.0E-15) wherein the possession of fewer A3 genes (e.g., placentals and rodents) is associated with lower accumulation levels, and a higher number of A3 genes (e.g., simiformes and some chiropterans) is associated with higher accumulation levels.

Correlation of A3 Gene Amplification and Diversification with ERV Activity. We examined the association between ERV invasions and A3 gene family expansion. As shown in Fig. 5A and B, we found that the number of A3 Z domains was positively associated with the percentage of ERVs in mammalian genome (in Poisson regression, coefficient = 0.14, P < 1.0E-15). Thus, species in which a greater proportion of the genome is composed of ERVs tend to have a higher number of A3 genes. Exceptions occur in the rodent family Muridae, as well as in 2 other species, hedgehog (Erinaceus europaeus) and opossum (M. domestica). In all of these outlier species, a large proportion of the genome is composed of ERV sequences, but relatively few or no A3 genes appear to be present (SI Appendix, Fig. S8A). As might be
expected, ERVs in these outlier species exhibited lower accumulation levels of G-to-A mutations overall (Fig. 5B). In addition, many of the ERVs identified in these species are relatively young (SI Appendix, Fig. S8 B–D) indicating that they derive from recent genome colonization events and have been incorporated into the germline without encountering A3-mediated mutation.

To investigate the association of A3 gene family expansion with ERV activity, we focused on primates because the evolutionary history of primate ERVs has been explored in depth and is relatively well characterized. We assessed the age of ERV invasions in each species using a genomic distance-based method and found that ERVs prominently invaded in the common ancestors of...
Simiiformes (including Hominioidea, Old World monkeys, and New World monkeys) around 50 million years ago (Fig. 5, C, Left). In contrast, ancestors of prosimians (including Lemurs, Lorisoids, and Tarsiers) did not experience prominent ERV invasion in this period. Furthermore, simians encoded higher numbers of A3 genes than prosimians (except for O. garnetti), suggesting that A3 gene amplification occurred early in the divergence of simian species (Fig. 5, C, Middle).

We investigated the timing of the formation of the double-domain A3G gene (i.e., A3G gene with A3Z2-A3Z1 structure) using the Ensembl gene database (www.ensembl.org). We found that simian primates encoded the double-domain (A3Z2-A3Z1) A3G gene, whereas prosimians did not, suggesting that the emergence of double-domain A3G genes also occurred during this period (Fig. 5, C, Right). Absence of a double-domain A3G gene in prosimians is supported by the finding that no A3Z2-A3Z1 genetic structures were observed in prosimian genomes (Fig. 3A). Overall, the timing of A3 gene amplification and diversification in primates was highly concordant with the timing of the prominent ERV invasions.

**Discussion**

Mammalian A3 family genes possess potent antiviral activities and are thought to have diversified during their evolution to allow targeting of a broader range of viruses (8, 12-14). ERVs provide a rich fossil record for retroviruses, enabling unique insights into the long-term coevolutionary interactions between retroviruses and their hosts. In the present study, we used the ERV fossil record to explore the coevolutionary history of A3 genes and ERVs.

When examining the ERV fossil record, it is vital to keep in mind that it is necessarily an incomplete record of retrovirus evolution. The vast majority of ERV sequences are fixed in the gene pool of host species, but since 1) fixation of any novel allele is extremely unlikely in the absence of strong selection and 2) most ERV insertions are likely to be selectively neutral at the time of insertion, it is reasonable to assume that the fixed ERVs we observe in the genomes of contemporary species represent a tiny subset of all of the ERVs that colonized their ancestors genomes. Furthermore, the ERV fossil record is presumably heavily biased toward retrovirus lineages that target germline cells, and there may have been many ancestral retrovirus lineages that never generated germline copies. Nonetheless, the fixed ERVs that are found in contemporary genomes are a unique source of retrospective information about the ancestral interactions between retroviruses and their hosts. Furthermore, because A3 genes restrict retrovirus replication via DNA editing, ERV sequences can contain genomic signatures that reveal information about their interactions with this particular group of restriction factors.

We show a strong positive correlation between A3 Z copy number and the extent to which G-to-A mutations have accumulated in ERV sequences (Fig. 4E). This finding reinforces the previously proposed concept (15, 16, 19, 21) that the accumulation of G-to-A mutations in ERVs reflects the antiviral activity of A3 proteins. We further show that mammalian species that have accumulated more ERVs (measured as a proportion of their ERV genome) tend to encode higher numbers of A3 genes and A3Z2-A3Z1 double-domain A3G genes.
Fig. 3. Genomic location of A3 genes. (A) Genomic order of the AID/APOBEC Z domains within the canonical A3 gene locus, which is sandwiched by CBX6 and CBX7 genes. Mammalian genomes in which CBX6 and CBX7 genes were detected in the same scaffold were only analyzed. The arrows indicate the direction of respective loci. (B) Bubble plot of the number of A3 Z domains in mammals. The number of the A3 Z domains in the whole genome (x axis) and that within the canonical A3 gene locus (y axis) in each mammal are plotted. Dot size is proportional to the number of species. (C) Genomic locations of A3 Z domains in S. boliviensis, A. nancymaae, and O. garnetti. A3 Z domains within 100 kb of each other were clustered. An asterisk denotes the A3 cluster corresponding to the canonical A3 gene locus. The arrows indicate the direction of respective loci. Pseudogenized sequences are indicated with an X. The sequences indicated by double daggers are intronless sequences and correspond to those described in SI Appendix, Fig. S5 A. The association between the genomic location of A3 genes and pseudogenization. The labels “in” and “out” denote the numbers of A3 Z domains located inside or outside the canonical A3 gene locus, respectively. Results for S. boliviensis, A. nancymaae, and O. garnetti are shown. Odds ratio and P value, calculated with Fisher’s exact test, are shown.

Most A3 genes are encoded in the canonical A3 locus and have been amplified by tandem gene duplication (Fig. 3 A and B). However, we also detected duplicated A3 genes outside this region in 3 primate species (S. boliviensis, A. nancymaae, and O. garnetti) (Fig. 3 B and C and SI Appendix, Fig. S5). All of these intronless A3G-like genes were amplified by retrotransposition. Furthermore, some are transcribed and may be functional (SI Appendix, Fig. S5).

A3 genes have been amplified in multiple lineages of mammals, but in addition, many A3 genes have been lost or pseudogenized (Fig. 1 C and SI Appendix, Fig. S3). For example, the A3Z1 gene was lost in Rodentia, and the A3Z3 gene was lost in Strepsirrhini and Microchiroptera. These findings might be attributed to genotoxic potential of these A3 genes: uncontrolled A3 expression can be harmful, and exogenous expression of human A3A (A3Z1 ortholog) in cell cultures triggers cytotoxic effects (48–50). Similarly, the aberrant expression of some human A3 proteins, particularly A3A (S1, S2), A3B (A3Z2–A3Z1 ortholog) (S1–S4), and A3H (A3Z3 ortholog) (S5), can contribute to cancer development by inducing somatic G-to-A mutations in the human genome.

Figure 3 details the genomic locations and distribution of A3 genes across various mammalian species. It illustrates how A3 genes have been amplified in tandem duplications, lost, or pseudogenized across different species, providing insights into the evolutionary dynamics of these genes. The analysis suggests a strong evolutionary pressure on the A3 gene family, with both amplification and loss occurring in various lineages.
Fig. 4. Signatures of A3 activity in ERV sequences and its association with A3 amplification. (A) Proportions of ERV sequences in the genomes of mammalian species. For proportions of LINE, SINE, and DNA transposon sequences, see SI Appendix, Fig. S6. (B) Strand bias scores of G-to-A mutation rates in human TEs (log2-transformed). The strand bias score is calculated as the G-to-A mutation rate ratio between the positive and negative strands. Dots indicate the strand bias score with respect to the strand bias score, the top 25 ERV subfamilies with respect to the variation (i.e., coefficient of variation) among the 4 G-to-A mutation sites (GA, GG, and GC) are shown. (D) ERV copies presenting the G-to-A hypermutation signature. ERV copies with >1 log2-transformed strand bias score and <0.1 false discovery rate are indicated as red. (E) Association of the number of A3 Z domains with the accumulation level of G-to-A mutations in ERVs in mammals. The x axis indicates the number of intact A3 Z domains, and the y axis indicates the mean value of the log2-transformed strand bias scores among ERVs in the genome. Correlation coefficient and P value are calculated by Pearson’s correlation.

Unlike the A3Z1 and A3Z2 genes, A3Z3 is highly conserved in most mammals and is not amplified in most mammalian lineages. Exceptions occur in carnivores and some other species; however, almost all duplicated A3Z3 genes identified in these species were pseudogenized (SI Appendix, Fig. S4). Moreover, phylogenetic relationships and the pattern of the premature stop codon
positions (SI Appendix, Fig. S4) indicate that the duplication–
psudogenization events have happened twice independently
during carnivore evolution. These observations support that
while the A3Z3 gene is indispensable for the hosts, its dupli-
cation might be genotoxic.

A3 proteins can suppress retroviral replication in a G-to-A
mutation-independent fashion (e.g., inhibition of reverse tran-
scription) (56–59). We could not address this dimension of
ERV–A3 interaction because of the technical difficulty of
assessing the mutation-independent effect of A3 proteins on
retroviruses using only genomic information. It should also be
noted that the number of A3 genes counted in this study might
underestimate the true value because of relatively low resolution
of many whole genome sequences. Moreover, we particularly
focused on the numbers and sequences of the Z domain of AID/
APOBEC family genes, and we could not fully address whether
1) some 2 Z domains compose a double domain gene and 2)
there are splicing variants. Nevertheless, this is to our knowledge
the most comprehensive investigation of A3 gene evolution
performed to date.

Materials and Methods
Sequence Data. WGS assemblies and RNA-Seq data analyzed in this study are
summarized in Datasets S1 and S6, respectively. Mammalian TE sequences were
obtained using RepeatMasker (version open-4-0-9) (http://repeatmasker.org)
with Repbase RepeatMasker libraries (version 20181026) (60). RMBlab
was selected as the search engine, and RepeatMasker was run with the options “q
xsmall -a -species” where <species> denotes the species name of the
analyzed genome (Dataset S7).

Genome Screening. Similarity search-based screens of sequence databanks
were performed using the database-integrated genome-screening (DIGS)
tool (61) which provides a relational database framework for performing
systematic tBLASTn-based screening of WGS databanks (61). We used AID/
APOBEC polypeptide sequences of 5 species (human, mouse, cow, megabat,
and cat) as queries for DIGS (SI Appendix, Fig. S1 A–C and Dataset S2).
The resultant list of hits (i.e., sequences disclosing homology to AID/APOBEC
family genes) was filtered to remove short and low-similarity matches
(tBLASTn bit score < 50). In the DIGS hit sequences, a partial sequence region

Fig. 5. Association between A3 gene family expansion and ERV invasion. (A and B) Association of the number of A3 Z domains with the amount of ERV
insertions in the genome. Dots are colored according to the species taxa (A) or the accumulation level of G-to-A mutations in ERVs (B). The association was
 evaluated under the Poisson regression with log link function. (C) Temporal association of ERV invasion with A3 gene amplification in primates. (Left)
Amount of ERV insertions in each age category in distinct primate species. ERV insertion date was estimated based on the genetic distance of each ERV integrant from
the consensus sequence under the molecular clock assumption [2.2 × 10−9 mutations per site per year (68)]. (Middle) Number of intact A3 Z domains. (Right)
Schematic of the MSA of A3G (A3Z2-Z3Z1 type) gene. Sequences of A3G genes in primates recorded in the Ensembl gene database (http://www.ensembl.org)
were used. NA, not applicable (no available data).
sequences, which have extremely long external branches (i.e., standard-
