

## In silico identification and Bayesian phylogenetic analysis of multiple new mammalian kallikrein gene families

Marc B. Elliott<sup>a,b</sup>, David M. Irwin<sup>a,c</sup>, Eleftherios P. Diamandis<sup>a,b,\*</sup>

<sup>a</sup> Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada M5G 1L5

<sup>b</sup> Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, ON, Canada M5G 1X5

<sup>c</sup> Banting and Best Diabetes Centre, University of Toronto, Toronto, ON, Canada, M5G 2C4

Received 24 February 2006; accepted 2 June 2006

Available online 7 July 2006

### Abstract

Kallikrein gene families have been identified previously in genomes of the human, the mouse, and the rat, and individual kallikrein-like genes have been found in many more species. This study presents the in silico identification of kallikrein gene families in the recently sequenced genomes of four additional mammalian species, the chimpanzee, the dog, the pig, and the opossum. Phylogenies were constructed with gene sequences from all seven mammalian families, using Bayesian analysis, which clarified the evolutionary relationships between these genes. Individual gene sequences, as well as concatenated constructs of multiple sequences, were used. Fifteen kallikrein genes were located in the chimpanzee (*Pan troglodytes*) genome, while only 14 were identified in the canine (*Canis familiaris*) genome as no orthologue to human *KLK3* was found. Thirteen genes were identified from the pig (*Sus scrofa*) genome, which lacked homologues to *KLK2* and *KLK3*, and 11 genes, orthologous to human *KLK5* through *KLK15*, were found in the opossum (*Monodelphis domestica*) genome. No kallikrein genes were identified from the available genome sequences of the chicken (*Gallus gallus*) or African clawed frog (*Xenopus tropicalis*). Within the family of kallikreins several subfamilies were suggested by phylogenetic analysis. One consisted of *KLK4*, *KLK5*, and *KLK14*; another of *KLK9*, *KLK11*, and *KLK15*; a third of *KLK10* and *KLK12*; a fourth of *KLK6* and *KLK13*; and finally one of *KLK8* and the classical kallikreins (*KLK1*, *KLK2*, and *KLK3*). © 2006 Elsevier Inc. All rights reserved.

**Keywords:** Kallikreins; Phylogenetics; Bayesian analysis; Gene duplication; Gene families

Kallikrein genes are trypsin- or chymotrypsin-like serine peptidase genes that exist as large multigene families [1]. To date, families from three separate species have been found and characterized, the human (*Homo sapiens*), mouse (*Mus musculus*) [2], and rat (*Rattus norvegicus*) [3,4], and individual genes have been found in the mastomys (a large African rodent) [5], cynomolgus monkey [6], rhesus monkey [7], dog [8], guinea pig [9], macaque, orang-utan, chimpanzee, gorilla [10], and cat [11], as well as in the horse, cow, and pig [4]. Kallikrein proteins were originally identified in the 1930s by Werle and colleagues, who found high levels in pancreatic isolates [12]. Initially there were thought to be two kallikrein genes: plasma kallikrein and tissue kallikrein [13]. During the 1980s, two

additional members of the tissue kallikrein family were identified: *KLK2*, dubbed human glandular kallikrein 2, and *KLK3*, or prostate-specific antigen (PSA) [14–16]. Together these three genes, *KLK1–3*, have been dubbed the “classical kallikreins.” In recent years, further work has expanded the number of tissue kallikreins in the human genome to 15 [17] and mapped them to a contiguous cluster spanning 400 kb on human chromosome 19q13.4 [18]. The human kallikrein locus is flanked by two non-kallikrein genes: the testicular acid phosphatase gene (*ACPT*) and the sialic acid-binding immunoglobulin-type lectin type 9 (*SIGLEC9*) gene. Under the current nomenclature, human kallikrein genes are designated by the symbol “*KLK*,” while human kallikrein proteins are designated by the symbol “hK” [19]. The term “tissue kallikrein” has also ceased to refer specifically to an enzyme with the ability to cleave a kininogen and release a kinin. Instead, it is now used to refer to any one of a family of gene

\* Corresponding author. Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, ON, Canada M5G 1X5. Fax: +1 416 586 8628.  
E-mail address: [ediamandis@mtnsinai.on.ca](mailto:ediamandis@mtnsinai.on.ca) (E.P. Diamandis).

products that share a highly conserved gene and protein structure as well as a significant degree of amino acid sequence identity. Indeed, many of these genes encode proteins with no appreciable kininogenase activity whatsoever, and the true in vivo functions of these proteins remain largely unelucidated.

To date, all kallikrein families found share the common trait that all their member genes are present in a single locus, as a contiguous cluster. In addition, all kallikrein genes found share four additional conserved features. All the genes have five coding exons encoding a putative serine peptidase. The codons encoding the catalytic triad residues are located at conserved positions; the histidine codon is near the end of coding exon 2, the aspartic acid codon is in the middle of coding exon 3, the catalytic serine codon is at the beginning of coding exon 5. Finally, intron phases (the point in a codon at which the intron is inserted) are fully conserved, at I–II–I–0. These five conserved characteristics, genomic organization, number of exons, exon

size, intron phase, and encoding a putative peptidase, have been set out as defining features of kallikrein genes [19]. An overview of the human kallikrein genes can be found in Fig. 1.

Numerous studies have suggested that human kallikreins may be useful as cancer biomarkers, particularly for hormone-dependent malignancies, such as ovarian, prostate, testicular, and breast carcinomas [20–23].

Phylogenetic analysis is a powerful tool for addressing many different evolutionary questions. Parsimony and maximum likelihood (ML) methods are the most established means of generating phylogenies (for a review, see [24]). Unfortunately, these methods are often time-consuming and grow in complexity at an exponential pace as more sequences are added to an analysis [25]. Bayesian methods are a more recent development in phylogenetic inference [26–30] and are closely related to ML methods. Bayesian methods are considerably faster than ML methods (up to 80%), especially for large datasets with more

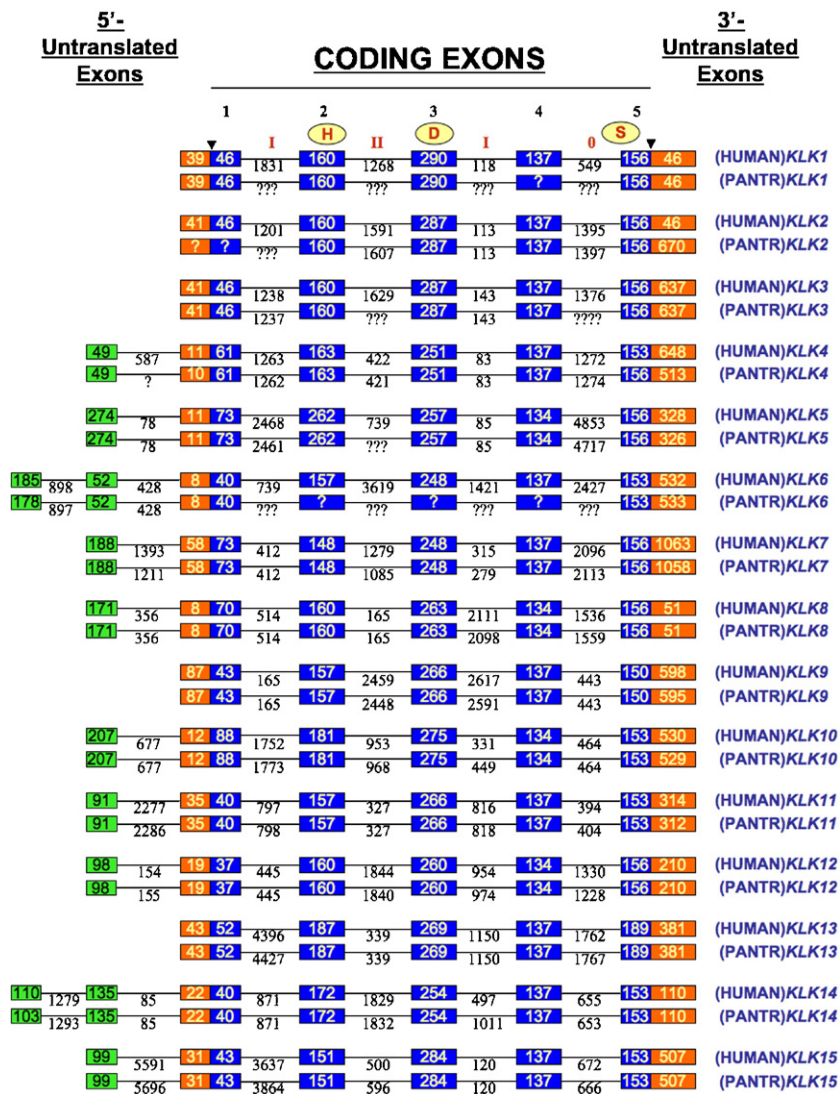


Fig. 1. Human and chimpanzee kallikrein genes. Human genes are indicated by the prefix (HUMAN), chimpanzee genes by (PANTR). Numbers indicate length in base pairs. Blue boxes indicate coding sequence. Green and orange boxes indicate untranslated exons/regions. The H, D, and S above the genes indicate the positions of the three catalytic residues in all genes. Roman numerals above introns indicate intron phase. Triangles indicate start and stop codons. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

complex models [31], and seem to give essentially the same results [32]. Few studies have been done to date on the evolution of the known kallikrein families. Two previous studies have been published [4,33], which have partially resolved some of the relationships between genes, but have left many questions unanswered, and focused on the large number of classical kallikrein paralogues that exist in the rodent lineages, which were not included in this study.

With the recent increase in whole-genome sequencing projects, large quantities of sequence data have become available

from multiple organisms. In this study, in silico gene identification methods were used to identify putative kallikrein homologues from publicly accessible databases. After manual annotation, Bayesian analysis methods were used to generate a phylogeny comprising the new kallikrein sequences alongside sequences from previously known families. In addition, gene sequences were concatenated together to create “supergenes”, which were then used to create phylogenies. It was hoped that these concatenated sequences would help to resolve any areas of low resolution in the basal regions of the phylogeny. While this

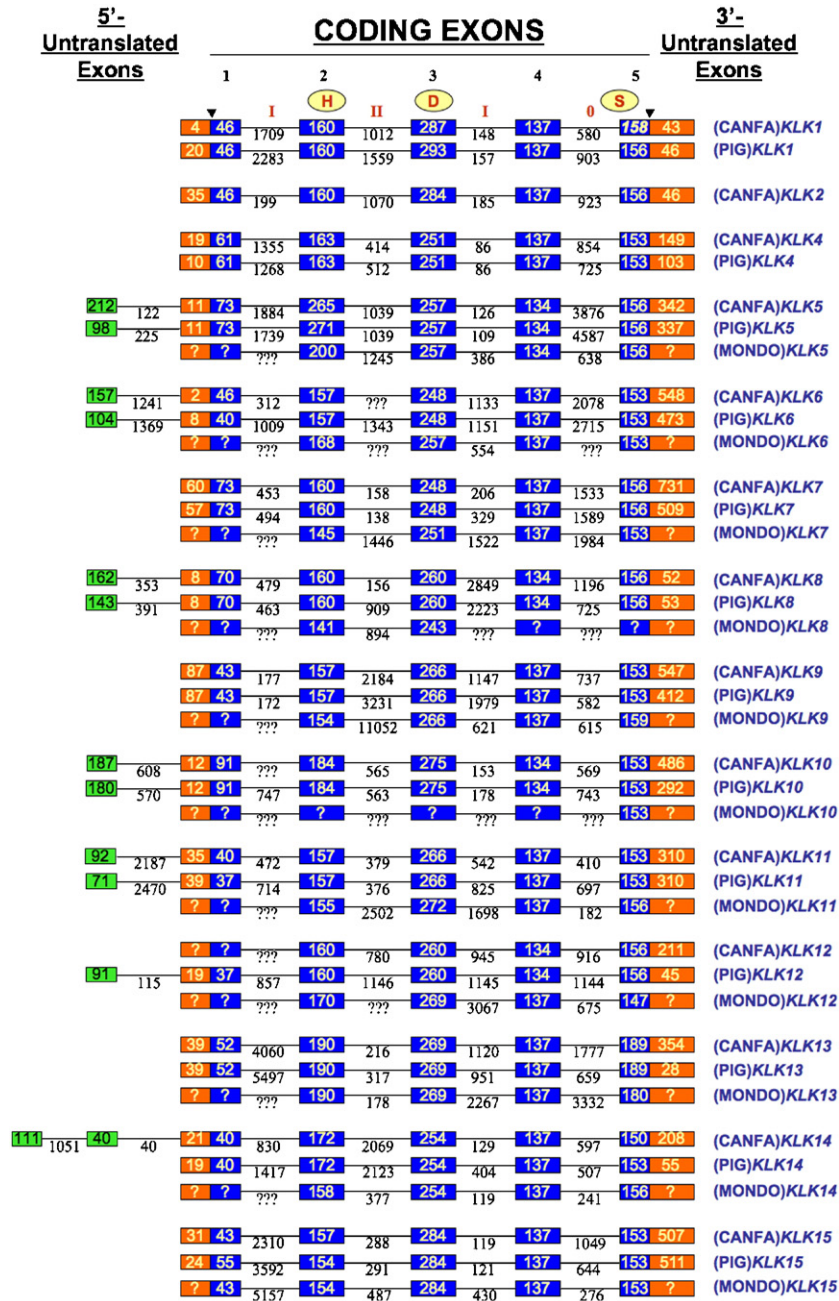


Fig. 2. Dog, pig, and opossum kallikrein genes. Dog genes are indicated by the prefix (CANFA), pig genes by (PIG), and opossum genes by (MONDO). Numbers indicate length in base pairs. Blue boxes indicate coding sequence. Green and orange boxes indicate untranslated exons/regions. The H, D, and S above the genes indicate the positions of the three catalytic residues in all genes. Roman numerals above introns indicate intron phase. Introns of indeterminate length due to genomic gaps are indicated by triple question marks. Unknown or indeterminate exon length is indicated by a single question mark. Triangles indicate start and stop codons. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



approach was devised independently in this case, it has been explored in the past [34]. These analyses were done to elucidate the evolution of the kallikrein gene family further and to confirm the identification of the new sequences identified.

## Results

### New kallikrein genes

Due to their strong identity with the human genes, the chimpanzee kallikreins are presented alongside their human orthologues in Fig. 1. Putative genes from the remaining three species in this study are presented together in Fig. 2. These arrangements were made to facilitate comparison between orthologous genes. Full listings of the genomic coordinates of all putative genes identified in this study are available as supplementary material, as are GenBank accession numbers for all previously characterized genes used.

A diagrammatic summary of the three known kallikrein families, along with the putative homologues identified in this study, can be found in Fig. 3.

### *Pan troglodytes*

The chimpanzee genome was found to contain 15 kallikrein genes, in the same genomic arrangement and orientation as the human family. The genes were located in a contiguous cluster on chimpanzee chromosome 19 (of the National Center for

Biotechnology Information (NCBI) assembly) or chromosome 20 (of the University of California at Santa Cruz (UCSC) assembly). Some genes were not fully covered by the available genomic sequence, but all of the chimpanzee genes whose sequence was fully available showed conservation of the basic characteristics of kallikrein genes, as described above. Within the coding region, the chimpanzee genes are identical to the human genes in terms of exon sizes. The untranslated exons in the majority of cases also appear to be of the same size. The major differences between human and chimpanzee kallikreins are in the intron sizes, although due to gaps in the genomic assembly, many chimpanzee introns are of indeterminate size.

### *Canis familiaris*

The dog genome was found to contain 14 kallikrein genes, all located in a contiguous cluster on canine chromosome 1. An orthologue to each human gene could be identified, with the exception of *KLK2* and *KLK3*, which in the dog are represented by a single gene, canine arginine esterase [8], more recently dubbed *Canis familiaris* kallikrein 2, as it shares enzymatic specificity with human *KLK2*, but not PSA. It is represented in this study as (CANFA)*KLK2*. The genes of the canine kallikrein family each appear to encode a functioning kallikrein-like gene. Significant differences exist in the intron sizes between the dog genes and the families of other species, as well as in the sizes and presence of untranslated exons found.

### *Sus scrofa*

From the available data, the pig genome (Swine Genome Sequencing Project) was found to contain at least 13 kallikrein genes. The genes appear to be orthologues of human kallikreins 1 and 4 through 15. These genes did appear in a contiguous cluster, but their chromosomal location in the porcine genome was unknown. The pig genes found all possess the conserved characteristics associated with kallikrein genes. As with the other families, intron size was quite variable, as were the length and number of untranslated exons found.

### *Monodelphis domestica*

The opossum genome (Broad Institute, Cambridge, MA, USA) scaffolds were found to contain 11 complete and partial kallikrein genes, which appear to be orthologues of human kallikreins 5 through 15. The genes found were present as contiguous clusters on three genomic scaffolds (12316, 4456, and 12183), which had no chromosomal localization in the monodelphine genome. Those genes and fragments that have been found all follow the same pattern as kallikrein genes from other species, with genes arranged in clusters, encoding serine peptidases with exons of similar size. For only the (MONDO) *KLK15* gene could all five coding exons be identified. The first coding exon of the remaining genes was not found; thus for the remaining genes only four or fewer coding exons could be identified. For those introns that could be identified, the intron phase pattern of the kallikrein genes (I–II–I–0) appeared to be conserved. The shorter length of the second exon of (MONDO) *KLK5*, compared to other species, is caused by the extra bases being contained within the signal peptide, which is poorly

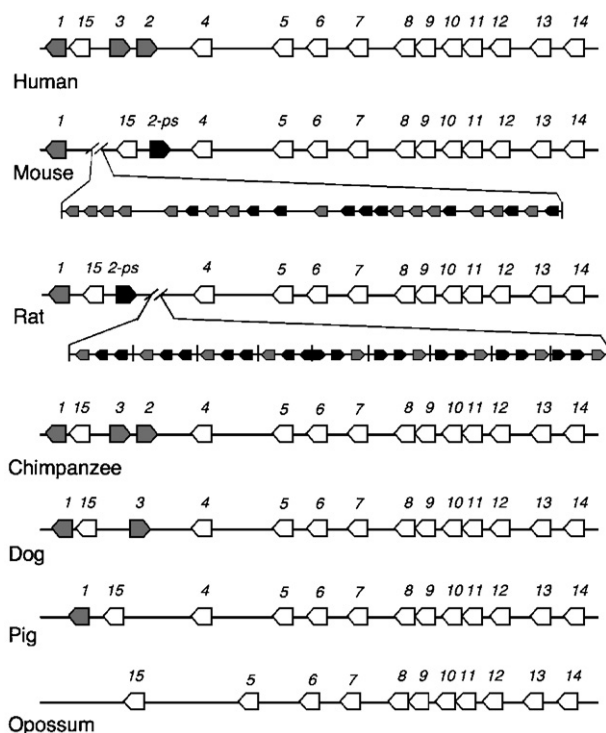


Fig. 3. Summary of known and new kallikrein families. Genes are indicated as arrows and have been arranged so as to indicate homology of vertical neighbors. Gray arrows indicate classical kallikreins, white arrows new kallikreins. Black arrows indicate pseudogenes, which, where named, are also given the suffix “-ps”. The multiple paralogues of *Klk1* present in the rodent lineages have been shown as insets below the rest of their families. Loci are drawn in approximate scale only.

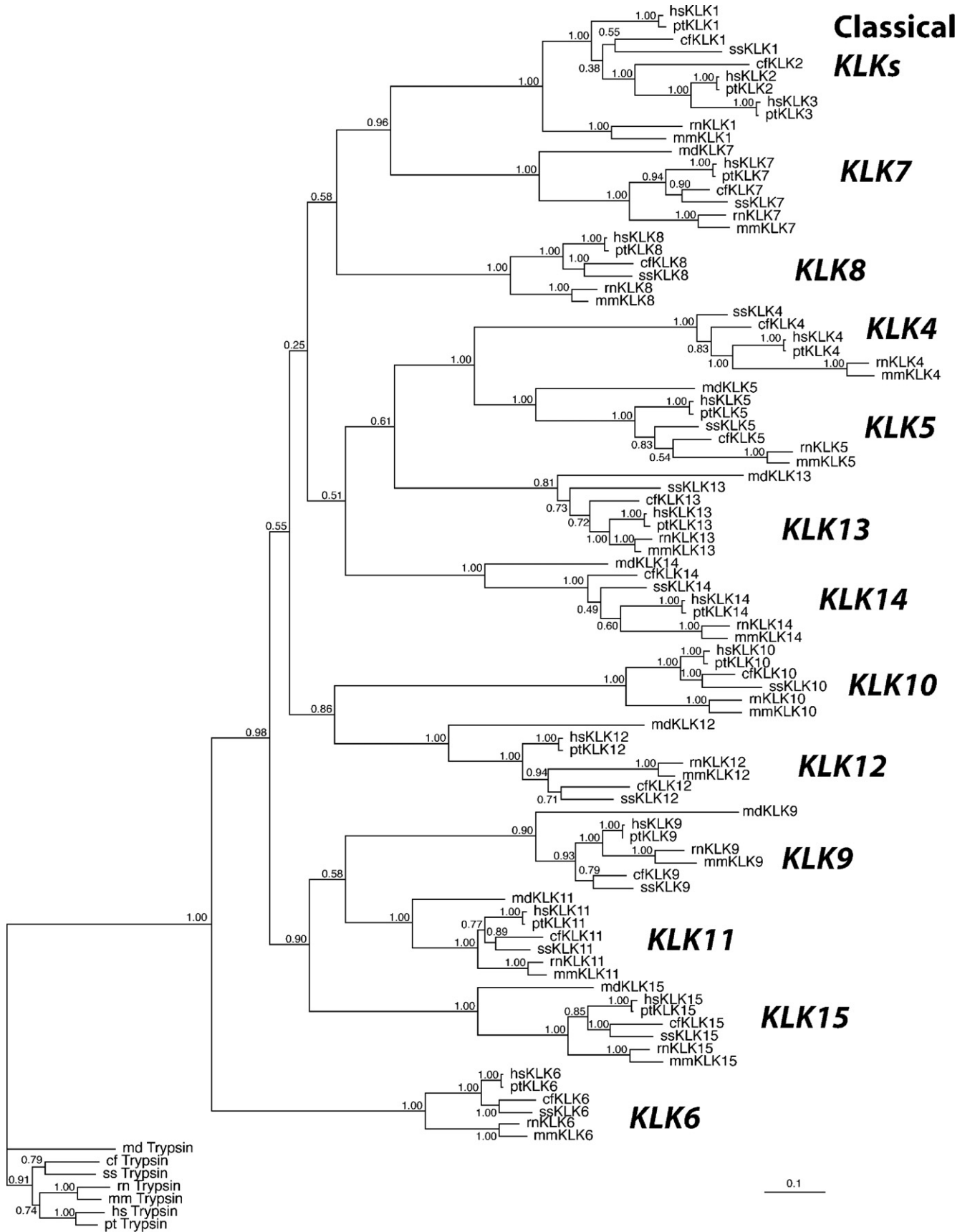


Fig. 4. Phylogenetic tree of individual kallikrein genes generated by Bayesian inference. Tree is constructed as a phylogram, with trypsin genes used as outgroup. Numbers behind nodes indicate support values for those nodes, up to a maximum of 1.00. The scale bar at the bottom right indicates a length of 0.1 nucleotide substitution per site.

conserved. No untranslated exons could be identified for any of the opossum genes.

#### *Gallus gallus*

Searches of the available chicken genome (International Chicken Genome Consortium) assembly using kallikrein gene and protein sequences yielded no hits that were more related to kallikreins than to other serine proteases and thus were not suitable for use in this analysis.

#### *Xenopus tropicalis*

The frog genome (DOE Joint Genome Institute) assemblies at UC Santa Cruz and NCBI were searched with kallikrein gene and protein sequences, as was the NCBI trace archive. Several contiguous clusters of serine peptidase genes were found, but when these groups were included in phylogenetic analyses, they were found not to be closely related to kallikreins.

#### Phylogenetic analysis

Two phylogenetic trees were generated in this study, one using individual kallikrein genes (Fig. 4), the other using concatenated gene sequences (Fig. 5). Both trees were largely congruent, but differed in the placement of two members of the kallikrein family, *KLK6* and *KLK7*. In the individual gene tree, *KLK7* was placed between *KLK8* and the classical kallikreins, whereas in the concatenated tree it was located between *KLK4/5* and *KLK14*. *KLK6* arose from the most basal node in the individual gene tree and clustered with *KLK13* as a sister group to the *KLK10/KLK12* group in the concatenated tree. Other than these two genes, however, the two trees are essentially identical. Both trees confidently grouped *KLK9* with *KLK11* and *KLK15* basal to them. Both trees also placed *KLK4* with *KLK5* and *KLK14* basal to them. *KLK10* was grouped with *KLK12*, and *KLK13* with them, and *KLK8* was grouped with the classical *KLKs*. Support values in the individual gene tree were weak

at certain nodes, but overall the tree was well supported. The tree made from concatenated genes had stronger support throughout, with only two nodes having support values of less than 0.83.

#### Discussion

Previous work on kallikrein families, along with rapidly expanding availability of genomic sequence data, has enabled us to examine the kallikrein content of multiple new genomes. Here we have identified, described, and conducted a phylogenetic analysis of 53 new kallikrein genes in four families. Coupled with previous data, this brings the number of known kallikrein families to 7 and the number of genes to an excess of 100. These genes can aid in our understanding of the evolution of the kallikrein family and could serve to develop model systems for studying gene functions in the future. Kallikreins have a known and established importance in the field of cancer diagnosis and prognosis, but further work is necessary to describe their physiological roles in a nonpathologic setting. The putative kallikrein families identified during this investigation seem to conform closely to previously established knowledge about these genes. All families had at most one copy of the *KLK4–KLK15* genes, with the greatest variability of duplications visible among the classical kallikreins, although none of the new families identified demonstrated anywhere near the number of duplications that have occurred in the rodent lineages. All the genes found encoded functional serine peptidase catalytic motifs, making it unlikely any of them had become pseudogenes, and all the established features used to define kallikrein genes were observed in the newly located genes.

The identification of kallikrein genes in a marsupial species is exciting, as it indicates that these genes are of very ancient origin and will likely be found as families in most, if not all, mammalian species. Various estimates place the marsupial–

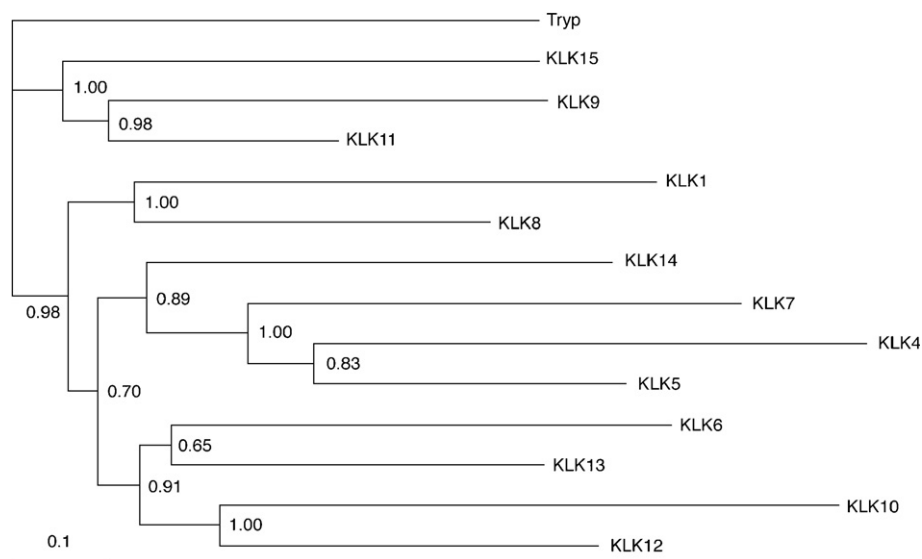


Fig. 5. Phylogenetic tree of concatenated kallikrein genes. Support at nodes is indicated by numbers beside them. Concatenated trypsin sequences were used as the outgroup. The scale bar at the bottom left indicates a length of 0.1 nucleotide substitution per site.

placental divergence at between 125 and 175 million years ago, and most of the duplications in the kallikrein family seem to have occurred before this. The exception to this seems to be the *KLK4/KLK5* duplication, which is not present in the monodelphine genome. Whether the duplication occurred after the marsupial lineage diverged or the (MONDO)*KLK4* gene was deleted cannot be determined from the phylogeny produced. Due to variable rates of substitution between kallikrein genes, and the possibility that (MONDO)*KLK5* may have evolved toward the same function as the other *KLK5* genes, no accurate chronology of the *KLK4/KLK5* duplication can be established. Evidence to address this question could come from the genome of the tamar wallaby (*Macropus eugenii*), which is currently being sequenced and will be the second marsupial genome available. That the kallikrein genes have survived with significant similarity in terms of sequence, organization, and number from this time also suggests that they fulfill important roles in normal cellular and physiological function. What also remains unknown is whether the marsupial lineage contains a homologue of *KLK1* or any classical kallikreins.

The additional sequence data obtained have enabled construction of a more detailed phylogeny of the kallikrein family than was previously available. That *KLK4* and *KLK5* are direct duplicates is a result supported by previous work, and the absence of *KLK4* in the opossum lends further support to this conclusion. In addition, *KLK4* seems to be one of the more rapidly evolving genes among the kallikreins, which suggests that it could be a more recent duplication still evolving toward a unique and specific function. The other relationship confirmed from previous work is the placement of *KLK9* and *KLK11* as sister taxa. That and the placement of *KLK10* and *KLK12* in our tree could suggest that a tandem duplication event took place, copying both genes at the same time. The tree made from individual genes seemed to indicate that *KLK6* was the member of the family that diverged earliest, but according to the concatenated tree it was the group of *KLK15*, *KLK9*, and *KLK11*, with *KLK15* basal to the other two. The other disagreement between the two trees regards the placement of the *KLK7* gene, which was alternately placed between *KLK8* and the classical *KLKs* or as a sister taxon to *KLK13*. Analysis of the concatenation method carried out by Gadagkar et al. [34] found that while concatenating sequences into supergene constructs in general yielded more accurate results, it also had the effect of exacerbating certain systematic biases. While the concatenated tree in this case was better supported than the individual gene tree, it has served to raise doubts about the proper placement of *KLK6* and *KLK7*.

The data collected in this investigation strongly indicate that kallikrein families exist in a large number of mammalian species. The phylogenetic trees generated here have served to clarify significantly the evolutionary history of the kallikrein family from what was known before. The discovery of credible and seemingly well-conserved homologues in the genome of the opossum, over 125 million years divergent from all other species in this study, is significant. Had kallikreins been found in either of the nonmammalian species examined, it would have shed much more light on the history of this gene family, but

further work will be needed to discover whether such genes exist. Interestingly, the available sequences do seem to indicate that the “classical kallikreins” are so merely because of their early discovery and that they in fact may be the most recent arrivals to the kallikrein family.

## Materials and methods

### Nomenclature

Due to the large number of genes found during this study, and the apparently widespread presence of kallikrein gene families in mammalian species, the preexisting nomenclature for kallikrein genes and proteins would have been insufficient. A new systematic nomenclature has been proposed by Lundwall et al. [35], which has been used here. In this system, use of the symbol “hK” or ones like it for proteins has been deprecated. Kallikrein genes are indicated as “*KLK*” and proteins as “*KLK*”. Where ambiguity would exist as to the species involved, the Swiss-Prot code is used (<http://www.expasy.ch/cgi-bin/speclist>) as a prefix in parentheses. Thus, under the new system, the human kallikrein 4 gene would be denoted (HUMAN)*KLK4*, and the canine kallikrein 6 protein as (CANFA)*KLK6*. The other symbols used here are (PANTR) for the chimpanzee, (PIG) for the pig, and (MONDO) for the opossum. When gene symbols are used here without a species prefix, they refer to all the homologues of that gene used in this analysis as a group.

### Identification of putative kallikrein homologues

The genomic sequences of the dog (version 1.0; Broad Institute), pig (GenBank contig AC149292), chimpanzee (Chimpanzee Sequencing Consortium, December 2003, build 1.1), opossum (BROAD 0.5; Broad Institute), chicken (GalGal 2; Genome Sequencing Center, Washington University School of Medicine at St. Louis, MO, USA), and frog (XenTro1; DOE Joint Genome Institute) were mined for kallikreins using a variety of computational tools. Significant use was made of the UCSC genome browser and BLAT alignment tool [36], as well as the BLAST and tBLASTn programs [37] and the Spidey mRNA-to-genomic alignment tool [38] from the NCBI Web site. Sequences from the previously identified kallikrein families in the human, rat, and mouse were used as probes. In brief, the available genomic data from each species (deposited in either the NCBI database or the UCSC genome browser) was searched with gene or protein sequences for various kallikreins, and any segments appearing to contain multiple hits were removed and searched in more detail using BLAST, tBLASTn, or Spidey and then manually annotated to identify splice sites as correctly as possible. The sequences of human *ACPT* and *SIGLEC9* were also used to aid in identifying possible kallikrein-containing regions. Once putative gene sequences were identified and annotated, they were aligned with known kallikrein sequences, and phylogenetic trees were generated to see if the sequences sorted themselves with the known kallikreins or in their own subtree. Sequences that sorted in their own subtree were rejected from the analysis.

### Gap-filling of new sequences

Due to the state of some of the genomic sequences used, there were at times gaps in some of the putative genes located. In such cases, efforts were made to fill these gaps so the sequences could be used in the phylogenetic analysis. Sequences were drawn from BAC clone, EST, and trace archive data, when it was available. In the case of two chimpanzee genes, (PANTR)*KLK1* and (PANTR)*KLK6*, human DNA sequence was used, as it was assumed that it would not differ significantly from the missing chimpanzee sequence.

### Sequence alignment and phylogenetic analysis

Multiple sequence alignment of identified gene sequences was performed with ClustalW 1.6 [39], using trypsin sequences from all seven species as



outgroups. To maintain the codon alignment, the sequences were translated and the proteins aligned, using the Gonnet protein weight matrix [40], a gap-opening penalty of 10, and a gap extension penalty of 0.1 for pair-wise alignment and 0.2 for multiple alignment. The resulting alignment was examined and corrected if it was deemed the program had made any small errors in aligning homologous residues. Next, any positions that contained gaps in more than 10% of the sequences or were excessively variable (more than 12 different amino acids between the various sequences) were removed from the alignment. The final result of this was an alignment of 98 sequences with 597 positions per sequence. The edited sequences were then used to construct the concatenated gene constructs, with species being arranged in the order human–chimpanzee–rat–mouse–dog–pig. Thus, in the concatenated alignments, no cross-species comparisons were made. The concatenated alignment was of 14 sequences, including outgroups, and had 3582 positions per sequence. Due to the lower number of available opossum sequences, they were not included in the concatenated constructs. In addition, as the *KLK2* and *KLK3* genes were not present in all the species used, those genes were not included in the concatenated gene constructs.

The final alignments were then used to construct phylogenetic trees by Bayesian analysis methods. To choose a model of evolution, and decide which additional parameters to use, the alignments were analyzed using ModelTest 3.7 [41] and PAUP\* 4.0b10 [42]. Following this, MrBayes 3.1.1 [27] was used to generate phylogenies for each alignment. For both alignments, the six-parameter General Time Reversible Model [43] was employed, with  $\gamma$ -distributed rates across sites and a proportion of sites allowed to be invariant. Two concurrent MCMC runs of 2,000,000 generations were used, each employing four progressively heated chains, with a temperature value of 0.2.

## References

- [1] G.M. Yousef, E.P. Diamandis, An overview of the kallikrein gene families in humans and other species: emerging candidate tumour markers, *Clin. Biochem.* 36 (2003) 443–452.
- [2] A.Y. Olsson, A. Lundwall, Organization and evolution of the glandular kallikrein locus in *Mus musculus*, *Biochem. Biophys. Res. Commun.* 299 (2002) 305–311.
- [3] D.R. Wines, J.M. Brady, D.B. Pritchett, J.L. Roberts, R.J. MacDonald, Organization and expression of the rat kallikrein gene family, *J. Biol. Chem.* 264 (1989) 7653–7662.
- [4] A.Y. Olsson, H. Lilja, A. Lundwall, Taxon-specific evolution of glandular kallikrein genes and identification of a progenitor of prostate-specific antigen, *Genomics* 84 (2004) 147–156.
- [5] M. Fahnestock, Characterization of kallikrein cDNAs from the African rodent *Mastomys*, *DNA Cell Biol.* 13 (1994) 293–300.
- [6] F.K. Lin, C.H. Lin, C.C. Chou, K. Chen, H.S. Lu, W. Bacheller, et al., Molecular cloning and sequence analysis of the monkey and human tissue kallikrein genes, *Biochim. Biophys. Acta* 1173 (1993) 325–328.
- [7] E.R. Gauthier, P. Chapdelaine, R.R. Tremblay, J.Y. Dube, Characterization of rhesus monkey prostate specific antigen cDNA, *Biochim. Biophys. Acta* 1174 (1993) 207–210.
- [8] P. Chapdelaine, E. Gauthier, M.A. Ho-Kim, L. Bissonnette, R.R. Tremblay, J.Y. Dube, Characterization and expression of the prostatic arginine esterase gene, a canine glandular kallikrein, *DNA Cell Biol.* 10 (1991) 49–59.
- [9] F. Fiedler, G. Betz, H. Hinz, F. Lottspeich, D.M. Raidoo, K.D. Bhoola, Not more than three tissue kallikreins identified from organs of the guinea pig, *Biol. Chem.* 380 (1999) 63–73.
- [10] J.F. Karr, J.A. Kantor, P.H. Hand, D.L. Eggenberger, J. Schlom, The presence of prostate-specific antigen-related genes in primates and the expression of recombinant human prostate-specific antigen in a transfected murine cell line, *Cancer Res.* 55 (1995) 2455–2462.
- [11] H. Fujimori, P.R. Levison, M. Schachter, Purification and partial characterization of cat pancreatic and urinary kallikreins—Comparison with other cat tissue kallikreins and related proteases, *Adv. Exp. Med. Biol.* 198 (Pt. A) (1986) 219–228.
- [12] E. Werle, Zur kenntnis des haushalts des kallikreins, *Biochem. Z.* 269 (1934) 415–434.
- [13] H.Z. Movat, E.G. Erdos, The plasma kallikrein–kinin system and its interrelationship with other components of blood, Bradykinin, Kallidin and Kallikrein, Springer-Verlag, Berlin, 1979, pp. 1–89.
- [14] K.W. Watt, P.J. Lee, M.T. Timkulu, W.P. Chan, R. Looor, Human prostate-specific antigen: structural and functional similarity with serine proteases, *Proc. Natl. Acad. Sci. USA* 83 (1986) 3166–3170.
- [15] A. Lundwall, H. Lilja, Molecular cloning of human prostate specific antigen cDNA, *FEBS Lett.* 214 (1987) 317–322.
- [16] L.J. Schedlich, B.H. Bennetts, B.J. Morris, Primary structure of a human glandular kallikrein gene, *DNA* 6 (1987) 429–437.
- [17] G.M. Yousef, E.P. Diamandis, The new human tissue kallikrein gene family: structure, function, and association to disease, *Endocr. Rev.* 22 (2001) 184–204.
- [18] G.M. Yousef, A. Chang, A. Scorilas, E.P. Diamandis, Genomic organization of the human kallikrein gene family on chromosome 19q13.3–q13.4, *Biochem. Biophys. Res. Commun.* 276 (2000) 125–133.
- [19] E.P. Diamandis, G.M. Yousef, J. Clements, L.K. Ashworth, S. Yoshida, T. Egelrud, et al., New nomenclature for the human tissue kallikrein gene family, *Clin. Chem.* 46 (2000) 1855–1858.
- [20] C.A. Borgoño, E.P. Diamandis, The emerging roles of human tissue kallikreins in cancer, *Nat. Rev. Cancer* 4 (2004) 876–890.
- [21] G.M. Yousef, E.P. Diamandis, Expanded human tissue kallikrein family—A novel panel of cancer biomarkers, *Tumour Biol.* 23 (2002) 185–192.
- [22] C.V. Obiezu, E.P. Diamandis, Human tissue kallikrein gene family: applications in cancer, *Cancer Lett.* 224 (2005) 1–22.
- [23] C.A. Borgoño, I.P. Michael, E.P. Diamandis, Human tissue kallikreins: physiologic roles and applications in cancer, *Mol. Cancer Res.* 2 (2004) 257–280.
- [24] P.O. Lewis, Phylogenetic systematics turns over a new leaf, *Trends Ecol. Evol.* 16 (2001) 30–37.
- [25] P.O. Lewis, D.L. Swofford, Back to the future: Bayesian inference arrives in phylogenetics, *Trends Ecol. Evol.* 16 (2001) 600–601.
- [26] B. Rannala, Z. Yang, Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference, *J. Mol. Evol.* 43 (1996) 304–311.
- [27] J.P. Huelsenbeck, F. Ronquist, MRBAYES: Bayesian inference of phylogenetic trees, *Bioinformatics* 17 (2001) 754–755.
- [28] B. Larget, D.L. Simon, Markov Chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees, *Mol. Biol. Evol.* 16 (1999) 750–759.
- [29] Z. Yang, B. Rannala, Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo method, *Mol. Biol. Evol.* 14 (1997) 717–724.
- [30] J.P. Huelsenbeck, F. Ronquist, R. Nielsen, J.P. Bollback, Bayesian inference of phylogeny and its impact on evolutionary biology, *Science* 294 (2001) 2310–2314.
- [31] M. Holder, P.O. Lewis, Phylogeny estimation: traditional and Bayesian approaches, *Nat. Rev. Genet.* 4 (2003) 275–284.
- [32] C.D. Kelly, Understanding mammalian evolution using Bayesian phylogenetic inference, *Mamm. Rev.* 35 (2005) 188–198.
- [33] A.Y. Olsson, C. Valtonen-Andre, H. Lilja, A. Lundwall, The evolution of the glandular kallikrein locus: identification of orthologs and pseudogenes in the cotton-top tamarin, *Gene* 343 (2004) 347–355.
- [34] S.R. Gadagkar, M.S. Rosenberg, S. Kumar, Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree, *J. Exp. Zool. B Mol. Dev. Evol.* 304 (2005) 64–74.
- [35] A. Lundwall, M. Blaber, J. Clements, Y. Courty, E. Diamandis, H. Lilja, et al., A comprehensive nomenclature for serine proteases with homology to the tissue kallikrein. *Biol. Chem.* (in press).
- [36] W.J. Kent, BLAT—The BLAST-like alignment tool, *Genome Res.* 12 (2002) 656–664.
- [37] S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, et al., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.* 25 (1997) 3389–3402.
- [38] S.J. Wheelan, D.M. Church, J.M. Ostell, Spidey: a tool for mRNA-to-genomic alignments, *Genome Res.* 11 (2001) 1952–1957.



- [39] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res.* 22 (1994) 4673–4680.
- [40] G.H. Gonnet, M.A. Cohen, S.A. Benner, Exhaustive matching of the entire protein sequence database, *Science* 256 (1992) 1443–1445.
- [41] D. Posada, K.A. Crandall, MODELTEST: testing the model of DNA substitution, *Bioinformatics* 14 (1998) 817–818.
- [42] D.L. Swofford, PAUP\*: Phylogenetic Analysis Using Parsimony (\*and Other Methods), version 4, Sinauer, Sunderland, MA; 2003.
- [43] F. Rodriguez, J.L. Oliver, A. Marin, J.R. Medina, The general stochastic model of nucleotide substitution, *J. Theor. Biol.* 142 (1990) 485–501.