



## Short Communication

## Comparative genomic analysis of eutherian kallikrein genes



Marko Premzl

Laboratory of Genomics, Centre of Animal Reproduction, 55 Heinzel St., Zagreb, Croatia

## ARTICLE INFO

## Article history:

Received 19 December 2016

Received in revised form 21 January 2017

Accepted 21 January 2017

Available online 3 February 2017

## Keywords:

Comparative genomic analysis

Gene annotations

Molecular evolution

Phylogenetic analysis

RRID:SCR\_014401

## ABSTRACT

The present study made attempts to update and revise eutherian kallikrein genes implicated in major physiological and pathological processes and in medical molecular diagnostics. Using eutherian comparative genomic analysis protocol and free available genomic sequence assemblies, the tests of reliability of eutherian public genomic sequences annotated most comprehensive curated third party data gene data set of eutherian kallikrein genes including 121 complete coding sequences among 335 potential coding sequences. The present analysis first described 13 major gene clusters of eutherian kallikrein genes, and explained their differential gene expansion patterns. One updated classification and nomenclature of eutherian kallikrein genes was proposed, as new framework of future experiments.

© 2017 The Author. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Brief communication

## 1.1. Introduction

The eutherian kallikrein genes were implicated in major physiological and pathological processes, as well as in medical molecular diagnostics [2, 11, 13, 23]. For example, the serum level of human prostate-specific antigen abundant in seminal plasma was used as molecular marker in diagnostics of prostate cancer. The comprehensive eutherian kallikrein gene data sets established major criteria of kallikrein gene annotations, including genomic localization within single genomic kallikrein locus region, common kallikrein gene features including 5 translated exons, as well as patterns of kallikrein protein amino acid sequence similarities including invariant catalytic triad amino acid sites His, Asp and Ser and distribution of common cysteine amino acid residues [2–4, 10, 11, 13, 15–17, 22, 23]. Specifically, the human kallikreins comprised 15 serine peptidases, including 3 classical kallikreins KLK1 or paradigmatic pancreatic kallikrein, KLK2 or glandular kallikrein and KLK3 or prostate-specific antigen, as well as KLK4–KLK15 kallikrein-related serine peptidases. Nevertheless, future updates and revisions of comprehensive eutherian gene data sets were expected, due to the incompleteness of public eutherian genomic sequence data sets [7, 8] and potential genomic sequence errors [9]. For example, the so-called lexicographical bias was described in some genomic sequence assembly programs [5] and phylogenetic analyses could be affected by potential genomic sequence errors [18]. Thus, the eutherian comparative genomic analysis protocol applicable in curation of major eutherian gene data sets was established as assistance in protection against potential genomic sequence errors [19–21]. The protocol

integrated gene annotations, phylogenetic analysis and protein molecular evolution analysis with new genomics and protein molecular evolution tests into one framework of eutherian gene descriptions. For example, the protocol revised, updated and published 9 major eutherian gene data sets, including 1172 complete coding sequences deposited in European Nucleotide Archive as curated third party data gene data sets. Using eutherian comparative genomic analysis protocol and public genomic sequence assemblies [1, 12, 14], the present study made attempts to update and revise eutherian kallikrein *KLN* genes.

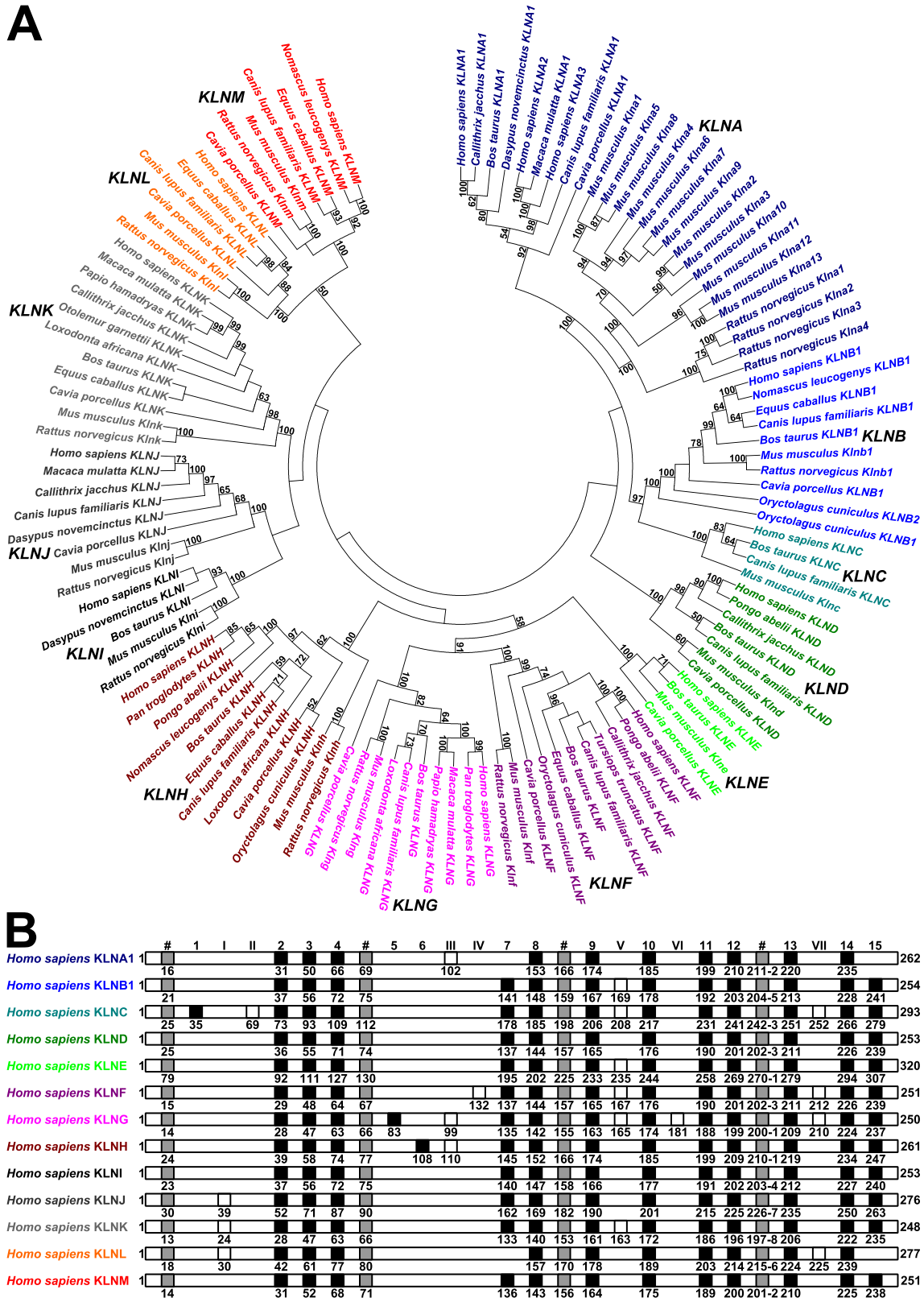
## 1.2. Gene annotations

The sequence alignment editor BioEdit 7.0.5.3 was used in analyses and manipulations of *KLN* nucleotide and protein sequences (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>). The public eutherian genomic sequence assemblies used in gene identifications of potential *KLN* coding sequences were downloaded from National Center for Biotechnology Information (NCBI) GenBank ([ftp://ftp.ncbi.nlm.nih.gov/genomes/genbank/vertebrate\\_mammalian/](ftp://ftp.ncbi.nlm.nih.gov/genomes/genbank/vertebrate_mammalian/)), as well as NCBI's BLAST programs (<ftp://ftp.ncbi.nlm.nih.gov/blast/>). In addition, the public eutherian genomic sequences were downloaded from Ensembl genome browser (<http://www.ensembl.org>). The analyses of *KLN* gene features included direct evidence of gene annotations deposited in NCBI's nr, est\_human, est\_mouse and est\_others databases (<https://www.ncbi.nlm.nih.gov>). The potential *KLN* coding sequences were tested using tests of reliability of eutherian public genomic sequences. The first test steps analysed nucleotide sequence coverages of each potential *KLN* coding sequence, using NCBI's BLAST programs and primary experimental genomic sequence reads deposited in NCBI's Trace Archive (<https://www.ncbi.nlm.nih.gov/Traces/trace.cgi>). Only if consensus read nucleotide sequence coverages were available for every nucleotide of each

E-mail address: [Marko.Premzl@alumni.anu.edu.au](mailto:Marko.Premzl@alumni.anu.edu.au).

potential *KLN* coding sequence, the potential *KLN* coding sequences were annotated as complete *KLN* coding sequences and used in analyses. Alternatively, the potential *KLN* coding sequences were described

as putative *KLN* coding sequences and not used in analyses. In addition, the complete *KLN* coding sequences were deposited in European Nucleotide Archive as eutherian third party data gene data set (<http://www.>



**Fig. 1.** (A) Phylogenetic analysis of eutherian kallikrein genes. The minimum evolution tree was calculated using maximum composite likelihood method. After 1000 replicates, the bootstrap estimates higher than 50% were shown. The major gene clusters *KLNA*–*KLNM* were indicated. (B) Eutherian kallikrein protein sequence alignment landmarks. Whereas the black squares labelled common cysteine amino acid residues (1–15), white squares labelled common N-glycosylation sites (I–VII) and grey squares labelled common exon-intron splice site amino acid sites (#). The numbers indicated numbers of amino acid residues.

ebi.ac.uk/ena/about/tpa-policy) [6]. The guidelines of human (<http://www.genenames.org/about/guidelines>) and mouse (<http://www.informatics.jax.org/mgihome/nomen/gene.shtml>) gene nomenclatures were used in updated *KLN* gene nomenclature and classification. The mVISTA's AVID program was used in multiple pairwise genomic sequence alignments of *KLN* genes using default settings (<http://genome.lbl.gov/vista/index.shtml>). In major cluster *KLNA* alignments, the cut-offs of detection of common genomic sequence regions in alignments with base sequence (*Homo sapiens KLNA1*) were 70% per 100 bp except 65% per 100 bp in pairwise alignments between base sequence and rodents. The common predicted *KLNA* promoter genomic sequence regions were aligned at nucleotide sequence level using ClustalW included in BioEdit 7.0.5.3, and manually corrected. Using BioEdit 7.0.5.3, the pairwise nucleotide sequence identities of common predicted *KLNA* promoter genomic sequence regions were calculated and used in statistical analyses, including calculations of average pairwise identities ( $\bar{a}$ ) and their average absolute deviations ( $\bar{a}_{ad}$ ), largest pairwise identities ( $a_{max}$ ) and smallest pairwise identities ( $a_{min}$ ) (Microsoft Office Excel). In other major cluster alignments (*KLNB-KLNM*), the cut-offs of detection of common genomic sequence regions in alignments with base sequences (*Homo sapiens*) were: 95% per 100 bp (*Pan troglodytes*), 90% per 100 bp (*Pongo abelii*, *Nomascus leucogenys*), 85% per 100 bp (*Macaca mulatta*, *Papio hamadryas*), 80% per 100 bp (*Callithrix jacchus*), 75% per 100 bp (*Otolemur garnettii*), 65% per 100 bp (rodents) or 70% per 100 bp in other pairwise alignments [19,20]. The RepeatMasker program version open-4.0.5 was used in detections and maskings of transposable elements in base sequences using default settings, except simple repeats and low complexity elements were not masked (sensitive mode, cross\_match version 1.080812, RepBase Update 20140131 and RM database version 20140131) (<http://www.repeatmasker.org/>). Therefore, among 335 potential coding sequences, the tests of reliability of eutherian public genomic sequences annotated most comprehensive curated third party data gene data set of eutherian *KLN* genes including 121 complete coding sequences (Fig. 1A) (Supplementary data file 1). Under accession numbers LT631550–LT631670, the eutherian *KLN* gene data set was deposited in European Nucleotide Archive (<https://www.ebi.ac.uk/ena/data/view/LT631550-LT631670>). First, the major cluster *KLNA* included 26 genes (*KLK1-KLK3*) (Supplementary data file 2A). The common predicted *KLNA* promoter genomic sequence region was annotated, having average pairwise nucleotide sequence identities  $\bar{a} = 0.74$  ( $a_{max} = 0.971$ ,  $a_{min} = 0.597$ ,  $\bar{a}_{ad} = 0.07$ ) (Supplementary data file 3). Whereas the major cluster *KLNB* included 10 genes (*KLK4*) (Supplementary data file 2B), major cluster *KLNC* included 4 genes (*KLK5*) (Supplementary data file 2C) and major gene cluster *KLND* included 7 genes (*KLK7*) (Supplementary data file 2D). The major cluster *KLNE* included 4 genes (*KLK15*) (Supplementary data file 2E), major gene cluster *KLNF* included 11 genes (*KLK9*) (Supplementary data file 2F), major cluster *KLNG* included 10 genes (*KLK11*) (Supplementary data file 2G) and major cluster *KLNH* included 12 genes (*KLK8*) (Supplementary data file 2H). Whereas the major cluster *KLNI* included 5 genes (*KLK6*) (Supplementary data file 2I), major cluster *KLNJ* included 8 genes (*KLK10*) (Supplementary data file 2J) and major gene cluster *KLNK* included 11 genes (*KLK12*) (Supplementary data file 2K). Finally, the major cluster *KLNL* included 6 genes (*KLK13*) (Supplementary data file 2L) and major cluster *KLNM* included 7 genes (*KLK14*) (Supplementary data file 2M). The present gene annotations confirmed major criteria of eutherian *KLN* gene annotations including genomic localization within single genomic kallikrein locus region and common *KLN* gene features including 5 translated exons, as well as different patterns of *KLN* differential gene expansions between eutherians [2–4,10,11,13,15,17,23]. Of note, whereas the human *KLN* gene annotations including 15 genes within kallikrein locus were likely complete, other kallikrein locus gene number estimates were subject to future updates and refinements due to incomplete genomic sequence assemblies and potential genomic sequence errors (Supplementary data file 1).

### 1.3. Phylogenetic analysis

Next, using BioEdit 7.0.5.3, the complete *KLN* coding sequences were translated and aligned at amino acid level using ClustalW included in BioEdit 7.0.5.3. After manual corrections of protein amino acid sequence alignments, *KLN* nucleotide sequence alignments were prepared accordingly. In *KLN* phylogenetic tree calculations, the MEGA 6.06 program was used (<http://www.megasoftware.net>), using neighbour-joining method (default settings, except gaps/missing data treatment = pairwise deletion), minimum evolution method (default settings, except gaps/missing data treatment = pairwise deletion), maximum parsimony method (default settings, except gaps/missing data treatment = use all sites) and unweighted pair group method with arithmetic mean method (default settings, except gaps/missing data = pairwise deletion). However, the maximum likelihood methods were not used in present analysis because their homogeneity and stationarity assumptions were not satisfied (data not shown). Using BioEdit 7.0.5.3, the pairwise nucleotide sequence identities of *KLN* nucleotide sequence alignments were calculated and used in statistical analyses, including calculations of average pairwise identities ( $\bar{a}$ ) and their average absolute deviations ( $\bar{a}_{ad}$ ), largest pairwise identities ( $a_{max}$ ) and smallest pairwise identities ( $a_{min}$ ) (Microsoft Office Excel). Therefore, the present phylogenetic analysis described 13 major gene clusters of eutherian *KLN* genes, using phylogenetic tree calculations (Fig. 1A) and calculations of pairwise nucleotide sequence identity patterns (Supplementary data file 4). For example, the 15 human *KLN* genes were distributed among 13 major gene clusters of eutherian *KLN* genes. The major eutherian *KLN* protein sequence alignment landmarks included protein primary structure features (Fig. 1B). Using different methods, very similar phylogenetic tree topologies were calculated (Fig. 1A). Of note, whereas the major gene clusters *KLNA* and *KLNB* included evidence of differential gene expansions respectively, major gene clusters *KLNC-KLNM* included orthologues respectively. There were major disagreements between present phylogenetic classification of eutherian *KLN* genes and analyses of Yousef et al. [23], Elliott et al. [3], Fernando et al. [4], Lundwall and Brattsand [13], Lawrence et al. [11], Pavlopoulou et al. [17], Marques et al. [15], Clements et al. [2] and Koumandou and Scorilas [10]. Specifically, the major gene cluster *KLNA* first integrated differentially expanded human *KLK1-KLK3* genes with differentially expanded rodent *KLN* genes. However, the clustering of major gene cluster *KLNB* (*KLK4*) with major gene cluster *KLNC* (*KLK5*) agreed with analyses of Yousef et al. [23], Elliott et al. [3], Fernando et al. [4], Pavlopoulou et al. [17] and Koumandou and Scorilas [10]. Yet, the clustering of major gene clusters *KLNE* (*KLK15*), *KLNF* (*KLK9*) and *KLNG* (*KLK11*) agreed with analyses of Yousef et al. [23], Elliott et al. [3], Fernando et al. [4], but disagreed with analyses of Pavlopoulou et al. [17] and Koumandou and Scorilas [10]. Finally, the clustering of major gene cluster *KLNL* (*KLK13*) with major gene cluster *KLNM* (*KLK14*) disagreed with analyses of Yousef et al. [23], Elliott et al. [3], Fernando et al. [4], Pavlopoulou et al. [17] and Koumandou and Scorilas [10]. The present phylogenetic classification of eutherian *KLN* genes was confirmed by calculations of pairwise nucleotide sequence identity patterns among 13 major gene clusters (Supplementary data file 4). First, the complete *KLN* nucleotide sequence alignments of 121 close eutherian homologues included average pairwise nucleotide sequence identity  $\bar{a} = 0.473$  ( $a_{max} = 0.994$ ,  $a_{min} = 0.266$ ,  $\bar{a}_{ad} = 0.066$ ). Among 13 eutherian *KLN* major gene clusters respectively, there were nucleotide sequence identity patterns of close eutherian orthologues and paralogues (*KLNA* and *KLNB*) and typical eutherian orthologues (*KLNC-KLNM*). In comparisons between eutherian *KLN* major gene clusters, there were nucleotide sequence identity patterns of very close eutherian homologues in comparisons between major gene clusters *KLNF*, *KLNG* and *KLNH*, in comparisons between major gene clusters *KLNE* and *KLNG*, as well as in comparisons between major gene clusters *KLNI* and *KLNM*. Finally, there were nucleotide sequence identity patterns of close eutherian homologues in other comparisons between major gene clusters.

#### 1.4. Protein molecular evolution analysis

The tests of protein molecular evolution included complete *KLN* nucleotide sequence alignments of 121 close eutherian homologues (Supplementary data file 5). The MEGA 6.06 program was used in calculations of *KLN* codon usage statistics. The relative synonymous codon usage statistics ( $R$ ) were determined as ratios between observed and expected amino acid codon counts. There were 22 not preferable amino acid codons ( $R \leq 0.7$ ), including TTT (0.47), TTA (0.1), TTG (0.46), CTT (0.55), CTA (0.32), ATT (0.61), ATA (0.37), GTT (0.53), GTA (0.22), TCG (0.37), CCG (0.46), ACT (0.61), ACG (0.48), GCG (0.43), TAT (0.47), CAT (0.5), CAA (0.48), AAT (0.61), GAT (0.65), GAA (0.56), CGT (0.49) and AGT (0.45). Using protein and nucleotide sequence alignments, the reference human *KLNA2* (prostate-specific antigen) protein sequence amino acid sites were then described as invariant amino acid sites (invariant alignment positions), forward amino acid sites (variant alignment positions) not including amino acid codons with  $R \leq 0.7$  or compensatory amino acid sites (variant alignment positions including amino acid codons with  $R \leq 0.7$ ). In *KLN* protein primary structures, the SignalP 4.1 server was used in predictions of N-terminal signal peptides using default settings (<http://www.cbs.dtu.dk/services/SignalP/>). Therefore, the eutherian *KLN* proteins were described as extracellular peptidases, including proenzyme or zymogen protein activation after cleavage of N-terminal signal peptides [11,16,23] (Supplementary data file 6). The major protein sequence alignment landmarks comprised mature eutherian *KLN* protein primary structure features, including common cysteine amino acid residues and N-glycosylation sites, as well as common exon-intron splice site amino acid sites (Fig. 1B). For example, there were 10–13 cysteine amino acid residues common to eutherian *KLN* major protein clusters respectively, including 10 invariant cysteine amino acid residues [2,23] (Supplementary data file 6). In addition, the eutherian *KLN* major protein clusters included 0–4 common N-glycosylation sites respectively [2,16,22]. Finally, the tests of protein molecular evolution integrated patterns of nucleotide sequence similarities with protein primary structures using present protein and nucleotide sequence alignments [19–21] (Supplementary data file 5). Specifically, among 261 reference human *KLNA2* amino acid residues, the present analysis described 27 invariant amino acid sites including 10 invariant cysteine amino acid residues, as well as 6 forward amino acid sites (Supplementary data file 6). For example, the invariant amino acid sites included catalytic triad amino acid sites His, Asp and Ser (H65, D120 and S213 in reference human *KLNA2* protein primary structure) [2,10,16,17,23]. In contrast, the analysis of Clements et al. [2] described 39 invariant amino acid sites in human *KLN* protein primary structures. Finally, there was one amino acid site cluster comprising overrepresented invariant and forward amino acid sites (C209–C219 in reference human *KLNA2* protein primary structure). For example, the amino acid site cluster included 2 invariant cysteine amino acid residues, catalytic triad amino acid site Ser, as well as amino acid motif GDSGGP common to serine peptidases [10], [17] and [23].

## 2. Conclusions

In conclusion, the most comprehensive curated third party data gene data set of eutherian *KLN* genes first annotated 13 major gene clusters *KLNA-KLNM* and explained their differential gene expansion patterns. As new framework of future experiments, the gene annotations, phylogenetic analysis and protein molecular evolution analysis integrated with new genomics and protein molecular evolution tests proposed updated classification and nomenclature of eutherian *KLN* genes.

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.ymgmr.2017.01.009>.

## Acknowledgements

The author would like to thank publisher on discounted article publication fees, and manuscript reviewers on their comments and suggestions.

## References

- [1] R.W. Blakesley, N.F. Hansen, J.C. Mullikin, P.J. Thomas, J.C. McDowell, B. Maskeri, A.C. Young, B. Benjamin, S.Y. Brooks, B.I. Coleman, J. Gupta, S.L. Ho, E.M. Karlins, Q.L. Maduro, S. Stantripop, C. Tsurgeon, J.L. Vogt, M.A. Walker, C.A. Masiello, X. Guan, N.I.S.C. Comparative Sequencing Program, G.G. Bouffard, E.D. Green, An intermediate grade of finished genomic sequence suitable for comparative analyses, *Genome Res.* 14 (2004) 2235–2244.
- [2] J.A. Clements, J.D. Hooper, Y. Dong, The human tissue kallikrein and kallikrein-related peptidase family, in: N.D. Rawlings, G.S. Salvesen (Eds.), *Handbook of Proteolytic Enzymes*, third ed. Academic Press, Oxford 2013, pp. 2747–2756.
- [3] M.B. Elliott, D.M. Irwin, E.P. Diamandis, In silico identification and Bayesian phylogenetic analysis of multiple new mammalian kallikrein gene families, *Genomics* 88 (2006) 591–599.
- [4] S.C. Fernando, F.Z. Najjar, X. Guo, L. Zhou, Y. Fu, R.D. Geisert, B.A. Roe, U. DeSilva, Porcine kallikrein gene family: genomic structure, mapping, and differential expression analysis, *Genomics* 89 (2007) 429–438.
- [5] P. Gajer, M. Schatz, S.L. Salzberg, Automated correction of genome sequence errors, *Nucleic Acids Res.* 32 (2004) 562–569.
- [6] R. Gibson, B. Alako, C. Amid, A. Cerdeño-Tarraga, I. Cleland, N. Goodgame, P. Ten Hoopen, S. Jayathilaka, S. Kay, R. Leinonen, X. Liu, S. Pallreddy, N. Pakseresht, J. Rajan, M. Rosselló, N. Silvester, D. Smirnov, A.L. Toribio, D. Vaughan, V. Zalusin, G. Cochrane, Biocuration of functional annotation at the European nucleotide archive, *Nucleic Acids Res.* 44 (2016) D58–D66.
- [7] J. Harrow, A. Frankish, J.M. Gonzalez, E. Tapanari, M. Diekhans, F. Kokocinski, B.L. Aken, D. Barrell, A. Zadissa, S. Searle, I. Barnes, A. Bignell, V. Boychenko, T. Hunt, M. Kay, G. Mukherjee, J. Rajan, G. Despacio-Reyes, G. Saunders, C. Steward, R. Harte, M. Lin, C. Howald, A. Tanzer, T. Derrien, J. Chrast, N. Walters, S. Balasubramanian, B. Pei, M. Tress, J.M. Rodriguez, I. Ezkurdia, J. van Baren, M. Brent, D. Haussler, M. Kellis, A. Valencia, A. Reymond, M. Gerstein, R. Guigó, T.J. Hubbard, GENCODE: the reference human genome annotation for The ENCODE Project, *Genome Res.* 22 (2012) 1760–1774.
- [8] International Human Genome Sequencing Consortium, Initial sequencing and analysis of the human genome, *Nature* 409 (2001) 860–921.
- [9] International Human Genome Sequencing Consortium, Finishing the euchromatic sequence of the human genome, *Nature* 431 (2004) 931–945.
- [10] V.L. Koumandou, A. Scorilas, Evolution of the plasma and tissue kallikreins, and their alternative splicing isoforms, *PLoS One* 8 (2013), e68074.
- [11] M.G. Lawrence, J. Lai, J.A. Clements, Kallikreins on steroids: structure, function, and hormonal regulation of prostate-specific antigen and the extended kallikrein locus, *Endocr. Rev.* 31 (2010) 407–446.
- [12] K. Lindblad-Toh, M. Garber, O. Zuk, M.F. Lin, B.J. Parker, S. Washietl, P. Kheradpour, J. Ernst, G. Jordan, E. Mauceli, L.D. Ward, C.B. Lowe, A.K. Holloway, M. Clamp, S. Gnerre, J. Alföldi, K. Beal, J. Chang, H. Clawson, J. Cuff, F. Di Palma, S. Fitzgerald, P. Flicek, M. Guttman, M.J. Hubisz, D.B. Jaffe, I. Jungreis, W.J. Kent, D. Kostka, M. Lara, A.L. Martins, T. Massingham, I. Moltke, B.J. Raney, M.D. Rasmussen, J. Robinson, A. Stark, A.J. Vilella, J. Wen, X. Xie, M.C. Zody, Broad Institute Sequencing Platform, Whole Genome Assembly Team, J. Baldwin, T. Bloom, C.W. Chin, D. Heiman, R. Nicol, C. Nusbaum, S. Young, J. Wilkinson, K.C. Worley, C.L. Kovar, D.M. Muzny, R.A. Gibbs, Baylor College of Medicine Human Genome Sequencing Center Sequencing Team, A. Cree, H.H. Dihn, G. Fowler, S. Jhangiani, V. Joshi, S. Lee, L.R. Lewis, L.V. Nazareth, G. Okwuonu, J. Santibanez, W.C. Warren, E.R. Mardis, G.M. Weinstock, R.K. Wilson, Genome Institute at Washington University, K. Delehaunty, D. Dooling, C. Fronik, L. Fulton, B. Fulton, T. Graves, P. Minx, E. Sodergren, E. Birney, E.H. Margulies, J. Herrero, E.D. Green, D. Haussler, A. Siepel, N. Goldman, K.S. Pollard, J.S. Pedersen, E.S. Lander, M. Kellis, A high-resolution map of human evolutionary constraint using 29 mammals, *Nature* 478 (2011) 476–482.
- [13] A. Lundwall, M. Brattsand, Kallikrein-related peptidases, *Cell. Mol. Life Sci.* 65 (2008) 2019–2038.
- [14] E.H. Margulies, J.P. Vinson, N.I.S.C. Comparative Sequencing Program, W. Miller, D.B. Jaffe, K. Lindblad-Toh, J.L. Chang, E.D. Green, E.S. Lander, J.C. Mullikin, M. Clamp, An initial strategy for the systematic identification of functional elements in the human genome by low-redundancy comparative sequencing, *Proc. Natl. Acad. Sci. U. S. A.* 102 (2005) 4795–4800.
- [15] P.I. Marques, R. Bernardino, T. Fernandes, N.I.S.C. Comparative Sequencing Program, E.D. Green, B. Hurler, V. Quesada, S. Seixas, Birth-and-death of *KLK3* and *KLK2* in primates: evolution driven by reproductive biology, *Genome Biol. Evol.* 4 (2012) 1331–1338.
- [16] R. Ménez, S. Michel, B.H. Muller, M. Bossus, F. Ducancel, C. Jolivet-Reynaud, E.A. Stura, Crystal structure of a ternary complex between human prostate-specific antigen, its substrate acyl intermediate and an activating antibody, *J. Mol. Biol.* 376 (2008) 1021–1033.
- [17] A. Pavlopoulou, G. Pampalakis, I. Michalopoulos, G. Sotiropoulou, Evolutionary history of tissue kallikreins, *PLoS One* 5 (2010), e13781.
- [18] H. Philippe, H. Brinkmann, D.V. Lavrov, D.T. Littlewood, M. Manuel, G. Wörheide, D. Baurain, Resolving difficult phylogenetic questions: why more sequences are not enough, *PLoS Biol.* 9 (2011), e1000602.
- [19] M. Premzl, Comparative genomic analysis of eutherian globin genes, *Gene Rep* 5 (2016) 163–166.
- [20] M. Premzl, Comparative genomic analysis of eutherian tumor necrosis factor ligand genes, *Immunogenetics* 68 (2016) 125–132.
- [21] M. Premzl, Curated eutherian third party data gene data sets, *Data Brief* 6 (2016) 208–213.
- [22] E.A. Stura, B.H. Muller, M. Bossus, S. Michel, C. Jolivet-Reynaud, F. Ducancel, Crystal structure of human prostate-specific antigen in a sandwich antibody complex, *J. Mol. Biol.* 414 (2011) 530–544.
- [23] G.M. Yousef, M.B. Elliott, A.D. Kopolovic, E. Serry, E.P. Diamandis, Sequence and evolutionary analysis of the human trypsin subfamily of serine peptidases, *Biochim. Biophys. Acta* 1698 (2004) 77–86.