Biochemical and Biophysical Research Communications **276,** 125–133 (2000)
doi:10.1006/bbrc.2000.3448, available online at http://www.idealibrary.com on IDEAL®

Genomic Organization of the Human Kallikrein Gene Family on Chromosome 19q13.3–q13.4

George M. Yousef,*,† Albert Chang,*,† Andreas Scorilas,*,† and Eleftherios P. Diamandis*,†,[1]

**Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario, Canada; and*
†Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada

Received August 11, 2000

Kallikreins are a subgroup of serine proteases with diverse physiological functions. Recently, growing evidence indicates that many kallikrein genes are involved in malignancy. In rodents, kallikreins are encoded by a large multigene family, but in humans only three kallikreins were thought to exist. Based on the homology between the human and rodent kallikrein loci, we studied a 300 kb region of genomic sequences around the putative KLK1 gene locus on chromosome 19q13.3–q13.4. By using linear sequence information, restriction analysis, end sequencing, PCR and blotting techniques, as well as bioinformatic approaches, we were able to construct the first detailed map of the human kallikrein gene family. Comparative analysis of genes located in this area, provides strong evidence that the human kallikrein gene family locus on chromosome 19 is considerably larger than previously thought, containing at least fifteen genes. We have established, for the first time, the common structural features that apply to all members of the expanded kallikrein multigene family. Our map specifies the distance between genes to one base pair accuracy, the relative location, and the direction of transcription of all 15 genes. Determination of the true size of the kallikrein family in humans is important for our understanding of the contribution of the kallikreins to human biology and pathophysiology. © 2000 Academic Press

Key Words: **kallikreins; kallikrein-like genes; serine proteases; human kallikrein locus; gene mapping; KLK3; chromosome 19.**

Abbreviations used: KLK, kallikrein; KLK-L, kallikrein-like; PCR, polymerase chain reaction; PSA, prostate specific antigen; dNTPS, deoxynucleoside triphosphates; LLNL, Lawrence Livermore National Laboratory; RT, reverse transcription; EST, expressed sequence tag; TLSP, trypsin-like serine protease; HSCCE, human stratum corneum chymotryptic enzyme; HSCTE, human stratum corneum tryptic enzyme; TADG-14, tumor associated differentially expressed gene 14; NES1, normal epithelial cell-specific 1 gene; kb, kilobase; BAC, bacterial artificial chromosome; PAC, P1-derived artificial chromosome; cM, centi-Morgan; BSA, bovine serum albumin; SSC, saline sodium chloride; SDS, sodium dodecyl sulfate; LB, Luria-Bertani; RACE, rapid amplification of cDNA ends; FISH, fluorescence in-situ hybridization; YAC, yeast artificial chromosome.

[1] To whom correspondence and reprint requests should be addressed at Mount Sinai Hospital, Department of Pathology and Laboratory Medicine, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada. Fax: 416-586-8628. E-mail: ediamandis@mtsinai.on.ca.

Kallikreins are a subgroup of serine proteases found in glandular cells, neutrophils, and biological fluids (1). Kallikreins are divided into two main groups: tissue and plasma kallikreins (2). The two groups differ in their molecular weight, substrate specificity, and immunological characteristics. A single gene encodes for the plasma kallikrein enzyme (3), whereas tissue kallikreins belong to a multigene family which shows different patterns of tissue specific expression (4). Recently, growing evidence suggests that many kallikreins and kallikrein-like genes are associated with cancer (5).

In mouse and rat, kallikreins are encoded by large multigene families. In the mouse genome, at least 24 genes have been identified (6). Expression of 11 of these genes has been confirmed; the rest are presumed to be pseudogenes (7). A similar family of 15–20 kallikreins has been found in the rat genome (8) where at least 4 of these are known to be expressed (9). The human kallikrein gene family was thought to have only three members: the tissue (pancreatic-renal) kallikrein (KLK1, encoding for hK1) (10), the human glandular kallikrein (KLK2, encoding for hK2) (11), and prostate specific antigen (KLK3, encoding for PSA or hK3) (12, 13). Several early estimations of the size of this family in humans, using Southern blot analysis, predicted just 3–4 genes (11, 14–16). However, a Southern blot analysis using monkey cDNA suggested that it might contain as many as 19 genes (17). Recently, new kallikreins and kallikrein-like genes have been discovered (18–25). In addition, a number of newly discovered serine proteases were found to have significant homologies with the classical kallikrein genes. Also, a recent report indicated the presence of two putative

0006-291X/00 $35.00

TABLE 1

Nomenclature for Human Kallikreins

New gene symbol[1,2]	Previous gene symbol(s)	New protein symbol	Other protein names/symbols
KLK1	KLK1	hK1	Pancreatic/renal kallikrein, hPRK
KLK15	—	hK15	—
KLK3	KLK3	hK3	Prostate specific antigen, PSA
KLK2	KLK2	hK2	Human glandular kallikrein 1, hGK-1
KLK4	PRSS17, KLK-L1, KLK4	hK4	Prostase, KLK-L1 protein, EMSP1
KLK5	KLK-L2	hK5	KLK-L2 protein; HSCTE
KLK6	PRSS9	hK6	Zyme, Protease M, Neurosin
KLK7	PRSS6	hK7	HSCCE
KLK8	PRSS19	hK8	Neuropsin; Ovasin; TADG-14
KLK9	KLK-L3	hK9	KLK-L3 protein
KLK10	PRSSL1, NES1	hK10	NES1 protein
KLK11	PRSS20	hK11	TLSP; Hippostasin
KLK12	KLK-L5	hK12	KLK-L5 protein
KLK13	KLK-L4	hK13	KLK-L4 protein
KLK14	KLK-L6	hK14	KLK-L6 protein

[1] Genes are presented in their order on chromosome 19q13.3–q13.4 from centromere to telomere.
[2] Full gene names are given under non-standard abbreviations.

kallikrein-like gene fragments in the region 19q13.3–q13.4 (26). Based on these findings, we hypothesized that the human kallikrein multigene family might be larger than previously thought.

The mouse kallikrein genes are clustered in groups of up to 11 genes on chromosome 7 and the distance between the genes in the various clusters can be as small as 3–7 kb (6). All three established human kallikrein genes have been assigned to chromosome 19q13.2–19q13.4 and the distance between KLK3 and KLK2 has been estimated to be 12 Kb (27).

Based on mapping of the rodent kallikrein genes and the documented strong conservation between human chromosome 19q13.3–q13.4 and the seventeen loci in a 20-cM proximal part of mouse chromosome 7 (28, 29), we identified a candidate genomic region (flanking the presumed KLK1 locus) for further analysis of the human kallikrein locus.

In this paper, we describe the construction of the first detailed map of the human kallikrein gene family, defining the order of genes along the chromosomal region, the direction of transcription, and estimating the distances between genes with one base pair accuracy. We provide strong evidence that kallikreins comprise a large multigene family in humans, and elucidate some common structural features that are conserved in all kallikrein genes.

MATERIALS AND METHODS

DNA sequences on chromosome 19. We obtained genomic sequences from the web site of the Lawrence Livermore National Laboratory (LLNL) (http://www-bio.llnl.gov/genome/genome.html). These sequences encompass a region flanking the putative KLK1 gene locus on chromosome 19q13.3–q13.4. The data were in the form of separate contigs constructed from BAC or cosmid clones. These contig sequences were subjected to an EcoR1 restriction analysis using the "webcutter" computer program (available at: http://www.firstmarket.com/cutter/cut2.html), and the fragment lengths were calculated.

Identification of positive BAC genomic clones from a human genomic DNA library. Many kallikrein genes were discovered independently and given various names. Recently, an international group of investigators established official nomenclature for these genes, as described in detail elsewhere (5). In Table 1, we present the conventional and official symbols of these genes. In the text, we use only the official nomenclature.

Gene-specific primers were used to amplify unique DNA sequences for the KLK1, KLK6, KLK13, KLK14 and the Siglec 9 gene (please see below). The primers used and the PCR reaction conditions are summarized in Table 2. The PCR products were purified and then labeled with ^{32}P by the random primer method (30), and used as probes to screen the PCRI-11 human male BAC library, constructed by de Jong and associates (31). The library was spotted in duplicate on nylon membranes, for identification of positive clones. The filters were hybridized in 15% formamide, 500 mM Na_2HPO_4, 7% SDS, 1% BSA (w/v) at 65°C overnight, then washed sequentially with 2× SSC, 1× SSC, 0.2× SSC, containing 0.1% SDS at 65°C, and then exposed to X-ray film as described (30). Positive clones were obtained, plated on selective LB medium, and then a single colony was transferred into LB broth for overnight cultures. Three PAC clones positive for KLK10 were identified by a similar methodology as described elsewhere (20).

End sequencing of BAC clones. Purification of BAC and PAC DNA was done by a rapid alkaline lysis miniprep method, which is a modification of the standard Qiagen-Tip method. Positive clones were further confirmed by Southern blot analysis as described (30). The ends of these clones were sequenced using vector-specific primers with an automated DNA sequencer.

Gene-specific amplification of other genes from genomic DNA. According to the published sequence of the 15 kallikrein genes identified in this region, we designed gene-specific primers for each of these genes and developed PCR-based amplification protocols and used them to screen the BAC and PAC clones described above. PCR was carried out in a reaction mixture containing 1 μl of DNA (0.1–1.0 μg), 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 2 mM $MgCl_2$, 200 μM dNTPs (deoxynucleoside triphosphates), 100 ng of primers and 2.5 units of

TABLE 2

Primers Used for Genomic PCR Amplification

Gene	Primer name and sequence		GenBank Accession No.	Coordinates	Length of PCR product (bp)	Annealing temperature (°C)
KLK-L1	L1-F1:	TGA CCC GCT GTA CCA CCC CA	AF135023	3723–5272	1500	63
	L1-R1:	GAA TTC CTT CCG CAG GAT GT				
KLK-L2	L2-2:	GGG CCT GTG GTC TGC AAT GG	AF135028	11104–11541	438	66
	L2-1:	GGG TCA GAG CTG CAG AGA AG				
KLK-L3	L3-F1:	AAC ATC AGC ATC CTG GAG AA	AF135026	7324–8070	747	64
	L3-R1:	CTT CGG CCT CTG TTG GTC TT				
KLK-L4	L4-FG1:	ACC CTG CCC AGA TAT GAA TG	AF135024	8071–8711	641	64
	L4-R1:	TTT ATT GTG GGC CCT TCA AC				
KLK-L5	K5-FG1:	GGC ATG GTG GCA TGC TTC TG	AF135025	3181–3840	659	68
	L5-R1:	TTG GTG ATG CCC CAG CCT GA				
KLK-L6	L6-FG1:	GGA AGG AAG TGC CCC GAC CT	AF161221	7082–7520	437	67
	L6-R1:	CCT GGA CTC CTG GGT CTG CA				
HSCCE	HS-FG1:	GTG AAG AAA GTC AGG CTG CC	AF166330	2178–2688	510	65
	HS2:	TGC CAG CGC ACA GCA TGG AA				
TLSP	TLSP-FG1:	TTG GAT AGC AGG CTG GGC AC	AF164623	6161–6742	582	68
	TLSP2:	TTC CTG CAC GCT GGC ACA CA				
Neuropsin	1S:	GTC ACC CCG CCC CTG GAT T	AF095743	645–1152	537	61
	N-R1:	GCT GCC TTC AGA TCC CTC CC				
Zyme	PRM-F:	GAC CCT GAC ATG TGA CATCTA	AF149289	11956–12397	442	65
	PRM-G:	GCC ACT GCC TGA TGG AGACTG				
Siglec-like	BPL-FG1:	CCT CTG CCT CCT AGG TTC AA	AF135027	5827–6474	648	64
	BPL-R2:	AAA AGG GAG GGC ACA GTG TG				
KLK1	KLK1-A:	ATC CCT CCA TTC CCA TCT TT	L10038	4081–4384	304	65
	KLK1-B:	CAC ATA CAA TTC TCT GGT TC				
KLK2	KLK2-A:	AGT GAC ACT GTC TCA GAA TT	M18157	131–580	450	65
	KLK2-B:	CCC CCA TCT CAC GAG TGC AC				
PSA	E5-A:	GTC GGC TCT GGA GAC ATT TC	M27274	5558–5931	374	62
	E5-B:	AAC TGG GGA GGC TTG AGT C				
NES1	NES15′-4:	GCT TCC CTA CCG CTG TGC T	AF055481	3790–4328	539	62
	NES13′-4:	CAC TCT GGC AAG GGT CCT G				

Note. Primer locations are according to the GenBank submissions.

HotStar Taq polymerase (Qiagen, Valencia, CA) on an Eppendorf thermal cycler. The cycling conditions were 95°C for 15 min to activate the HotStar Taq polymerase, followed by 40 cycles of 94°C denaturation for 30 s, annealing for 30 s, and 72°C extension for 1 min, and a final extension at 72°C for 10 min. Equal amounts of PCR products were electrophoresed on 2% agarose gels and visualized by ethidium bromide staining. Primers and annealing temperatures used for the different reactions are summarized in Table 2.

Cloning and sequencing of the PCR products. Due to the high degree of homology between the genes in this genomic region, all primers were designed to be specific for each gene, annealing away from conserved regions. To further verify the identity of the PCR products, they were cloned into the pCR 2.1-TOPO vector (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. The inserts were sequenced from both directions using vector-specific primers, with an automated DNA sequencer.

Long PCR amplification. A pair of primers connecting the first exon of KLK1 (CTC CGA TCC CCA GAG CTG TA) and the centromeric end of the KLK3 genomic sequence (CTG GTG GTG CCA AGG ATC CA), were used to amplify the gap between KLK1 and KLK3. PCR was performed using the Expand Long Template PCR System (Roche Molecular Systems, Indianapolis, IN) with human genomic DNA as a template. We used a 50 μl reaction mixture containing 300 ng DNA, 20 mM Tris-HCl (pH 7.5), 100 mM KCl, 2.25 mM $MgCl_2$, 1 mM dithiothreitol (DTT), 0.1 mM EDTA, 0.5% (v/v) Tween 20, 0.5% (v/v) Nonidet P40, 500 μM dNTPs, 300 nM primers and 2.6 units of enzyme mix containing thermostable Taq and Pwo DNA polymerases, on a Perkin-Elmer 9600 thermal cycler. The cycling conditions were 94°C for 2 min, followed by 10 cycles of 94°C for 10 s, 63°C for 30 s, 68°C for 8 min, then 20 cycles of 94°C for 10 s, 63°C for 30 s, 68°C for 8 min (increasing by 20 s per subsequent cycle), in addition to a final extension step at 68°C for 7 min. The PCR products were electrophoresed on a 0.8% agarose gel.

Purification and sequencing of long PCR products. PCR products were purified with exonuclease I plus shrimp alkaline phosphatase treatment as described elsewhere (32). Briefly, 1 μl of exonuclease I and 1 μl of shrimp alkaline phosphatase were added to the PCR products, then the mixture was incubated at 37°C for 15 min, then at 85°C for 15 min. Excess oligonucleotides and dNTPs were removed by a spin dialysis method using Centricon 30 microconcentrators (Amicon, (30)). After purification, PCR products were directly sequenced using an automated DNA sequencer.

RESULTS

Construction of a Continuous Chromosomal Region Encompassing the Human Kallikrein Gene Locus

We obtained genomic sequences around the human kallikrein gene locus on chromosome 19q13.3–13.4 from the Lawrence Livermore National Laboratory. These sequences were in the form of separate contigs

TABLE 3

Results of BAC and PAC Library Screening with Gene-Specific Primers for All Kallikrein Genes

	PAC 43B1	PAC 121B1	PAC 86023	BAC 288H1	BAC 76F7	BAC 10I11	BAC 433D04	BAC 1N23	BAC 195D22
KLK1	−	−	−	−	−	−	−	−	−
PSA	−	−	−	+	+	−	−	−	−
KLK2	−	+	−	+	+	−	−	−	−
KLK-L1	−	+	−	+	+	−	−	−	−
KLK-L2	−	+	+	+	+	−	−	−	−
Zyme	−	+	+	+	+	−	−	−	−
HSCCE	−	+	+	+	+	+	−	−	−
Neuropsin	+	+	+	+	+	+	−	−	−
KLK-L3	+	+	+	+	+	+	−	−	−
NES1	+	+	+	−	−	+	−	−	−
TLSP	−	+	+	−	−	+	−	−	−
KLK-L5	−	−	+	−	−	+	−	−	−
KLK-L4	−	−	−	−	−	+	+	−	−
KLK-L6	−	−	−	−	−	+	+	−	−
Siglec	−	−	−	−	−	−	+	+	+

with different lengths. Comparison of the *Eco*R1 restriction maps of the contigs with the restriction maps of their parent clones (available from the LLNL web site) enabled us to align the contigs along the chromosomal region. Some contigs were overlapping, enabling us to construct a contiguous segment; however, few gaps were present.

A Blast search was done for the genomic sequence against the non-redundant GenBank database (33). This allowed us to define the precise location of two classical kallikrein genes; PSA (KLK3), KLK1 and KLK2. We also localized other newly discovered serine proteases which are homologous with the kallikrein genes, KLK6 (19), human stratum corneum chymotryptic enzyme (KLK7) (34), KLK8 (23), normal epithelial cell-specific 1 gene (KLK10) (20, 35), and trypsin-like serine protease (KLK11) (36, 37).

Three gaps were present in our genomic sequence. These gaps were closed as follows: (1) The margins of the first gap (G1) were found to contain the 5′ and 3′ ends of the KLK2 gene; this enabled us to fill the gap with the published genomic structure of the KLK2 gene (GenBank Accession No. M18157). (2) The margins of the third gap (G3) were found to contain the 3′ and 5′ ends of the KLK6 gene mRNA sequence; thus, a radiolabeled probe specific for the KLK6 gene was used to screen a human BAC library (see Materials and Methods) and two positive clones were obtained. Restriction digestion was performed, followed by Southern blotting and a fragment containing the KLK6 gene was obtained and sequenced, thus filling this gap (25). (3) The second gap (G2) was closed by performing a long PCR reaction using primers located at the margins of a BAC clone that covers this area and the BAC clone as a template (see below).

Identification of New Kallikrein Genes in the Human Kallikrein Locus

A number of computer programs were used to predict the presence of putative new genes within the contiguous genomic region and six new putative genes were predicted, as described elsewhere (38). The full structure of these genes was verified by using various methods, including EST sequencing and homology search, PCR, 3′ and 5′ RACE, as described elsewhere (24, 25, 39). In addition, a novel member of the Siglec gene family was identified at the telomeric end of the genomic sequence (GenBank Accession No. AF135027).

Relative Localization of the Kallikrein Genes

Verification of the accuracy of the constructed human kallikrein gene locus and determination of the relative localization of the kallikrein genes were performed as follows:

(A) Screening of human BAC and PAC libraries. A human BAC library was screened using radiolabeled probes specific to KLK1, KLK6, KLK13, KLK14, and the Siglec 9 gene. Six BAC clones which were positive to any of the probes were selected for further analysis. Three PAC clones, known to be positive for the KLK10 gene (20), were also included.

(B) PCR screening of the BAC and PAC clones. The above-mentioned clones were screened by PCR using gene-specific primers for each of the 15 kallikrein genes that were identified in this region (Table 2). The results of this screening are summarized in Table 3 and a representative gel is shown in Fig. 1. These data are consistent with our predicted order for the 15 genes along the physical map.

(C) End sequencing of BAC and PAC clones. Further confirmation of the relative location of these genes

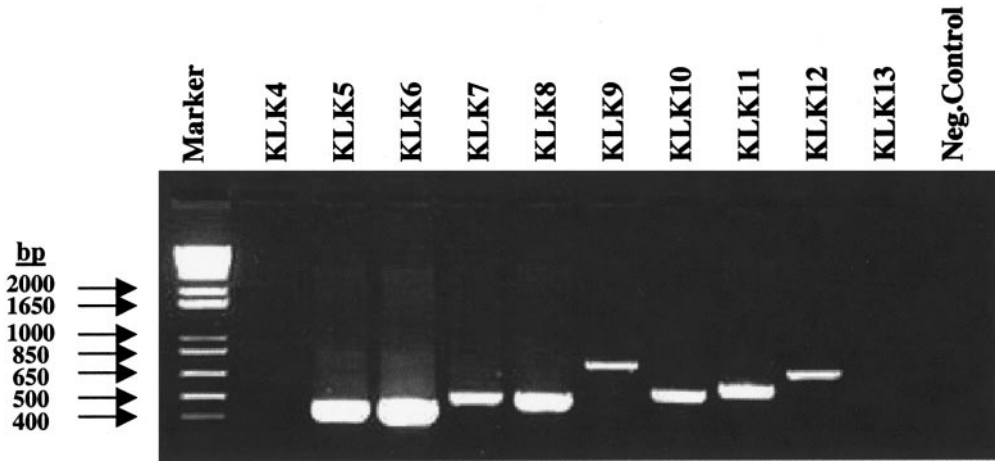

FIG. 1. PCR products for PAC clone 86023. The figure shows presence of 8 tandemly localized genes and absence of another two.

in relation to each other was obtained by end sequencing of the BAC and PAC clones. Each clone was sequenced from both ends, and the sequences obtained were aligned with the genomic sequence to identify the ends of each clone. Figure 2 shows a diagrammatic representation of genomic areas covered by each clone (horizontal lines), and the ends of each clone as determined by end sequencing (vertical lines).

The data of Fig. 2 and Table 3 are consistent with the predicted physical map of this region. The relative locations of individual genes in relation to each other were verified as follows (see Fig. 2 and Table 3): (1) The position of KLK1 in relation to KLK3, and the direction of transcription of both genes was verified by performing a long PCR reaction, with a forward primer from the reverse complementary sequence of the first exon of KLK1 and a reverse primer from the sequence adjacent to the KLK15 gene, using human genomic DNA as a template. The reverse primer was located 2 kb from the centromeric end of the contig, containing the KLK3 gene and the PCR product was a band of ~9 kb in length (data not shown). The PCR product was cloned and sequenced to verify the specificity of the reaction. This data confirms both the tandem localization of the KLK1 and KLK3 genes and the distance between them, as reported by others (~31 kb). (2) Using bioinformatics, we have further predicted the existence of a novel serine protease gene, mapping between KLK1 and KLK3. This gene was then cloned and sequenced and was found to transcribe in the same direction as KLK1. The gene is known as KLK15 (GenBank Accession No. AF242195). (3) BAC 288H1 is positive for both KLK3 and KLK2, and PAC 121B1 is positive for KLK2, but not KLK3 (Fig. 2), thus, KLK3 must be centromeric to KLK2, also in accordance with previous reports (26). (4) KLK4 and KLK2 are present in two different contigs, and the alignment of these contigs according to the *Eco*R1 restriction map indicates that KLK4 is telomeric to KLK2 (data not shown). (5) KLK5 must be telomeric to KLK4, because PAC clone 86023 is positive for KLK5 but not for KLK4. (6) The KLK6 gene sequence bridges the gap between two adjacent contigs, 26 and 27 of clone 85745 (numbers refer to contigs of the LLNL) so, KLK6 must be at the telomeric end of contig 26, and thus further downstream from KLK5. (7) BAC 10I11 is positive for KLK7 but not KLK6, and in addition, its end sequence falls within the first part of the KLK6 gene; thus, KLK6 must be centromeric to KLK7. (8) The KLK8 gene is telomeric to KLK7, as confirmed by the fact that both genes are located on one contiguous fragment that fits the *Eco*R1 map of this region. (9) By comparing telomeric sequences of clones 288H1, 76F7, 121B1 and 86023, we verified the order of genes telomeric to KLK8 as follows: KLK9, KLK10, KLK11, KLK-L5 and KLK13. (10) By obtaining additional sequences from the centromeric end of BAC 433D04 by primer walking, we confirmed that KLK13 is centromeric to KLK14. (11) Finally, we observed that BAC10I11 contains the telomeric group of this gene family, except for the novel Siglec 9 gene, indicating that this gene is the most telomeric. Siglec 9 is not a kallikrein or a serine protease and its location indicates the end of the kallikrein gene family and the start of another gene family (Siglecs).

In Fig. 3 we present the physical location of the 15 members of the human kallikrein gene family, around the chromosomal locus 19q13.3–q13.4. The lengths of the genes and between genes are given with 1 bp accuracy. These distances may change slightly in the future if more untranslated exons are found in the 3′ or 5′ ends of these genes. Additionally, the transcription initiation site of each gene has not as yet been accurately defined.

In Fig. 4 we present the genomic organization of the 15 genes, describing intron and exon sizes, start and stop codons, untranslated exons, intron phases and location of the amino acids of the catalytic triad.

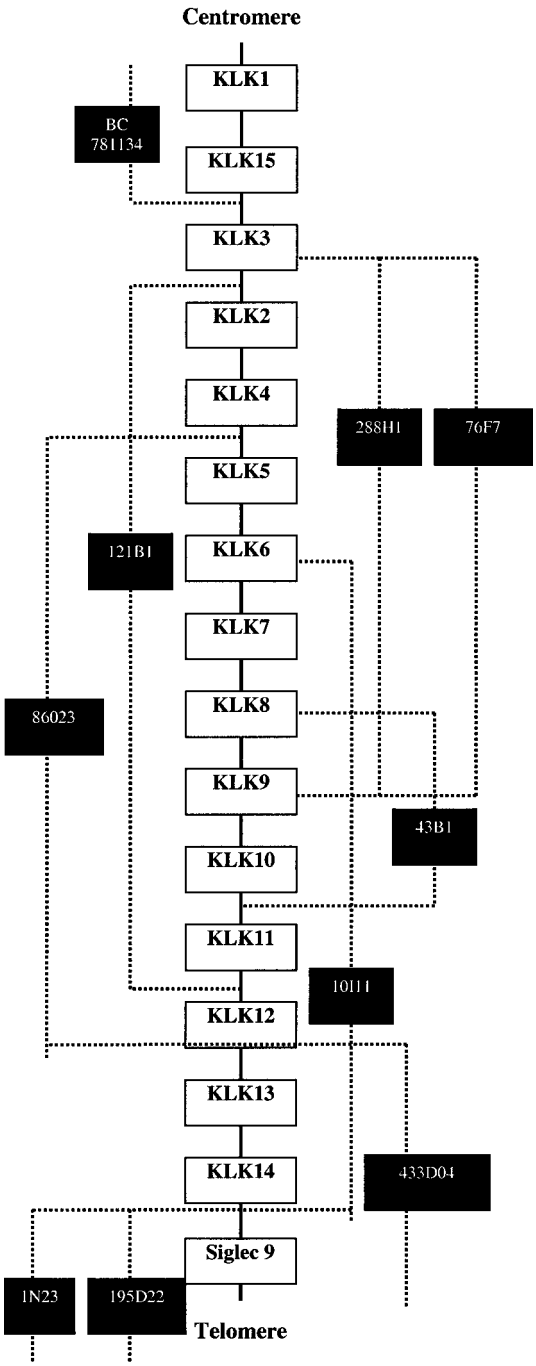

FIG. 2. Schematic presentation of the results of PCR and end-sequencing for the different BAC and PAC clones. The genes are shown, in order, in open boxes along a horizontal line and BAC and PAC clone IDs are inside solid bars. Horizontal dotted lines represent areas covered by the corresponding clones, and vertical dotted lines the locations of both ends of each clone, as determined by end-sequencing. Figure is not drawn to scale. For complete gene names and details, see Tables 1 and 3 and text.

DISCUSSION

Kallikreins are involved in diverse physiological processes through their serine protease enzymatic activity (2). Determination of the gene family size and cloning of all its members is important to our understanding of the contribution of the kallikreins to human biology.

In this report, we have precisely defined the human kallikrein gene locus and constructed the first detailed

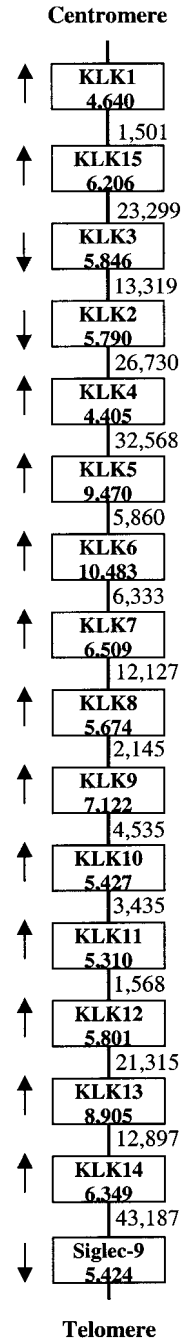

FIG. 3. An approximate 300 kb region of contiguous genomic sequence around chromosome 19q13.3–q13.4. The direction of transcription of each gene is illustrated by arrows. Boxes represent genes and contain the gene name and its genomic length in base pairs. Distances between genes are described in base pairs between boxes. Figure is not drawn to scale. For full gene names see abbreviations footnote and Table 1.

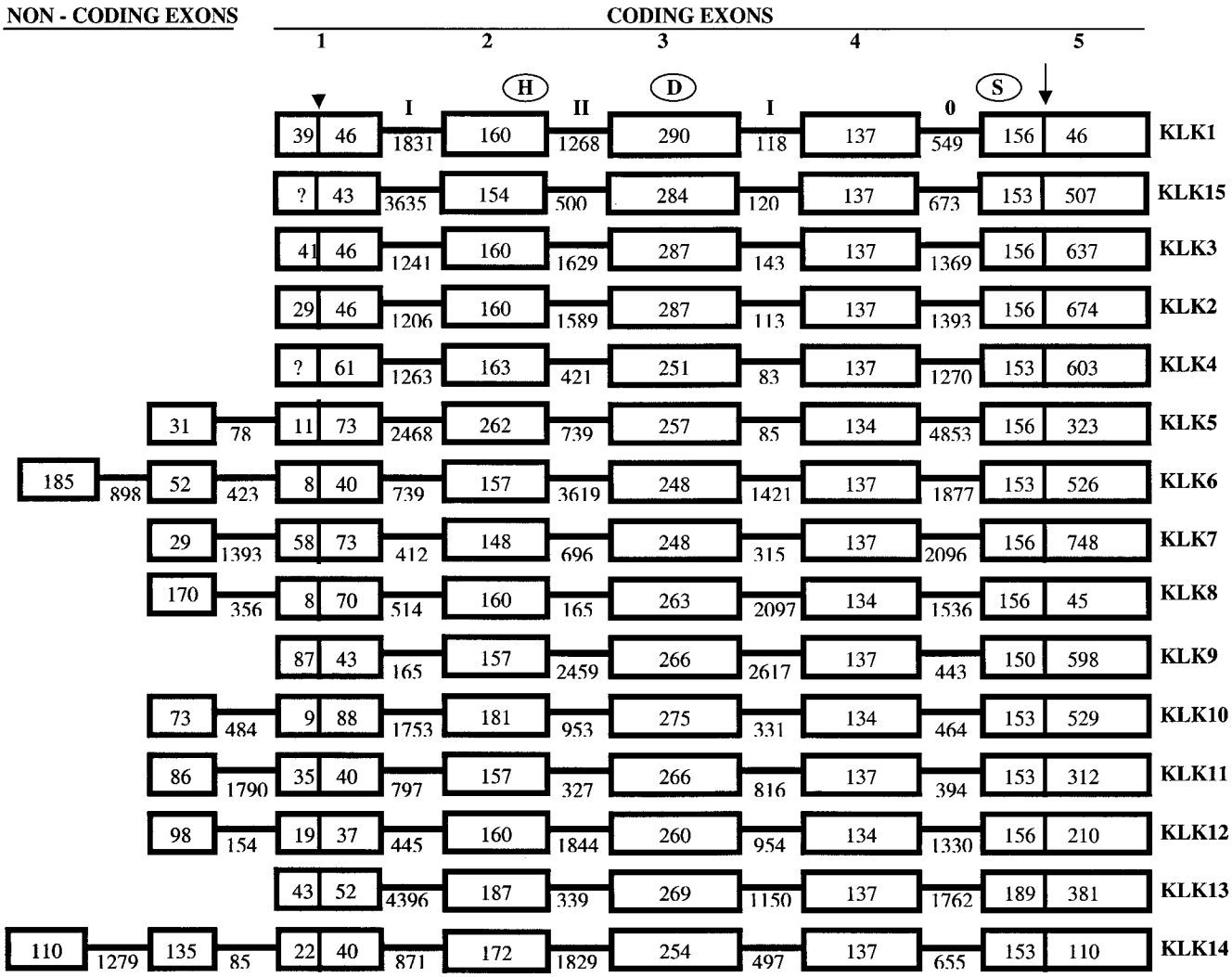

FIG. 4. Schematic diagram showing the comparison of the genomic structure of the fifteen genes found in the putative human kallikrein locus. Exons are shown by solid bars and introns by the connecting lines. Arrowhead marks the site of the start codon, and the arrow represents the stop codon. Letters above boxes indicate relative positions of the catalytic triad that was found to be conserved in all genes; H denotes histidine, D denotes aspartic acid, and S denotes serine. Roman numbers indicate intron phases. The intron phase refers to the location of the intron within the codon; I denotes that the intron occurs after the first nucleotide of the codon, II the intron occurs after the second nucleotide, 0 the intron occurs between codons. The intron phases are conserved in all genes. Numbers inside boxes indicate exon lengths in base pairs, and those below the connecting lines show the intron lengths. Question mark denotes that region length is unknown. Figure is not drawn to scale.

map describing the relative positions of the kallikreins and other kallikrein-like genes (Fig. 3). This map is consistent with previous reports on the localization of the classical kallikreins (26, 27), and the approximate mapping of some new kallikreins by radiation hybrid and FISH techniques (20, 21, 38). It should be noted, however, that the estimation of intervals between genes depends on the published sequences of the genes, and might change slightly in the future, since some kallikreins may have extra 5′ exon(s) that have not as yet been identified. This map is also directional; it indicates that KLK3 and KLK2 genes are transcribed in the same direction (centromere to telomere) and that the rest of the kallikrein-like genes are transcribed in the reverse direction. The possibility still exists that this locus is extended further, and that other kallikrein-like genes may be located upstream of KLK1, downstream from KLK14 or in between. The second possibility is more unlikely since the novel Siglec 9 gene identified, belongs to a multigene family which is localized telomeric to the kallikrein locus (our unpublished data).

In our study, all results of BAC and PAC end sequencing and screening were consistent with each other as well as with the *Eco*R1 restriction map of the region, with the exception of one PAC clone (43B1), which was positive for KLK10, KLK9, KLK8 and KLK1 genes and negative for all other genes. These results

were not matching with data from all other PAC and BAC clones, and this clone was excluded from the analysis.

Kallikreins were traditionally characterized by their ability to liberate lysyl-bradykinin (kallidin) from kininogen (2). More recently, however, a new, structural concept has emerged to describe kallikreins. From accumulated sequencing data, it is now clear that the mouse has many genes with high homology to kallikrein coding sequences (40, 41). The concept of a "kallikrein multigene family" proposed by Richards and co-workers, refers to these genes (42, 43). This definition is not based much on specific enzymatic function of the gene product, but more on its sequence homology and their close linkage on mouse chromosome 7. In humans, only KLK1 meets the functional definition of a kallikrein. Based on the newer definition, members of the kallikrein family include, not only the gene for the kallikrein enzyme, but also genes encoding other homologous proteases that are located in the same chromosomal region. Therefore, it is important to note the clear distinction between the enzyme kallikrein and a kallikrein or a kallikrein-like gene.

An early report proposed that the serine proteases could be classified into five different groups according to intron position (44). The established kallikreins (KLK1, KLK2, and KLK3) belong to a group in which: (1) The intron phases are always conserved (I-II-I-0) (see Fig. 4 for description of intron phases). (2) There is an intron just downstream from the codon for the active site histidine residue, (3) a second intron is present downstream from the exon containing the codon for the active site aspartic acid residue, and (4) a third intron is located just upstream from the exon containing the codon for the active site serine residue. Our results indicate the presence of some additional common structural features that are found in all kallikreins, as shown in Fig. 4: (1) All genes are formed of 5 coding exons and 4 intervening introns (with the possibility that some genes may have extra 5′ untranslated exon(s) (20, 23, 38). (2) The exon lengths are usually comparable. (3) These genes are clustered in the same chromosomal region, apparently without any intervening non kallikrein-like genes (Fig. 3). Thus, we can conclude that all the recently identified serine proteases that are present in this region could be considered members of the expanded human kallikrein multigene family.

Evidence is accumulating that many of the kallikreins and kallikrein-like genes that are clustered in the same chromosomal region are related to malignancy. KLK3 is the best marker for prostate cancer so far (13). A recent report provided evidence that KLK3 has antiangiogenic activity, and that this activity may result from its action as a serine protease. This study also suggested that other serine proteases and members of the kallikrein multigene family of enzymes should be evaluated for potential antiangiogenic action (45). Recent literature suggests that KLK2 could be a useful diagnostic marker for certain subtypes of prostate cancer (46). KLK10 is a tumor suppressor gene (47). KLK6 is differentially expressed in primary breast and ovarian tumors (18), and the human stratum corneum chymotryptic enzyme (KLK7) has been shown to be expressed at abnormally high levels in ovarian cancer (48). Another recently identified kallikrein-like gene, the tumor-associated differentially expressed gene-14 (TADG-14/Neurospin; KLK8) was found to be overexpressed in a subset of ovarian cancers (49). Also, KLK4 is speculated to be linked to prostate cancer (21).

It is possible that some of these kallikreins participate in pathways which are involved in tumor initiation, progression, or suppression. Identifying the relative positions of these genes in relation to one another may be important for their regulation. Many of these genes have already been found to be regulated by steroid hormones, as summarized elsewhere (5). We anticipate that the detailed description of the genomic organization of the human kallikrein gene family will facilitate further progress in the field and accelerate discoveries which may have practical clinical value.

ACKNOWLEDGMENTS

We thank Linda Ashworth for her assistance in using the sequencing data from the Lawrence Livermore National Laboratory and Dr. S. Scherer for advice and help in the screening of the genomic libraries.

REFERENCES

1. Bhoola, K. D., Figueroa, C. D., and Worthy, K. (1992) *Pharmacol. Rev.* **44,** 1–80.
2. Clements, J. (1997) *in* The Kinin System (Farmer, S., Ed.), pp. 71–97, Academic Press, New York.
3. Beaubien, G., Rosinski-Chupin, I., Mattei, M. G., Mbikay, M., Chretien, M., and Seidah, N. G. (1991) *Biochemistry* **30,** 1628–1635.
4. Mason, A. J., Evans, B. A., Cox, D. R., Shine, J., and Richards, R. I. (1983) *Nature* **303,** 300–307.
5. Diamandis, E. P., Yousef, G. M., Luo, L. Y., Magklara, A., and Obiezu, C. V. (2000) *Trends Endocrinol. Metab.* **11,** 54–60.
6. Evans, B. A., Drinkwater, C. C., and Richards, R. I. (1987) *J. Biol. Chem.* **262,** 8027–8034.
7. Drinkwater, C. C., Evans, B. A., and Richards, R. I. (1988) *Trends Biochem. Sci.* **13,** 169–172.
8. Ashley, P. L., and MacDonald, R. J. (1985) *Biochemistry* **24,** 4520–4527.
9. Gerald, W. L., Chao, J., and Chao, L. (1986) *Biochim. Biophys. Acta* **867,** 16–23.
10. Evans, B. A., Yun, Z. X., Close, J. A., Tregear, G. W., Kitamura, N., Nakanishi, S., Callen, D. F., Baker, E., Hyland, V. J., Sutherland, G. R., *et al.* (1988) *Biochemistry* **27,** 3124–3129.
11. Schedlich, L. J., Bennetts, B. H., and Morris, B. J. (1987) *DNA* **6,** 429–437.

12. Riegman, P. H., Vlietstra, R. J., van der Korput, J. A., Romijn, J. C., and Trapman, J. (1989) *Biochem. Biophys. Res. Commun.* **159,** 95–102.
13. Diamandis, E. P. (1998) *Trends Endocrinol. Metab.* **9,** 310–316.
14. Riegman, P. H., Vlietstra, R. J., Klaassen, P., van der Korput, J. A., Geurts van Kessel, A., Romijn, J. C., and Trapman, J. (1989) *FEBS Lett.* **247,** 123–126.
15. Fukushima, D., Kitamura, N., and Nakanishi, S. (1985) *Biochemistry* **24,** 8037–8043.
16. Baker, A. R., and Shine, J. (1985) *DNA* **4,** 445–450.
17. Murray, S. R., Chao, J., Lin, F. K., and Chao, L. (1990) *J. Cardiovasc. Pharmacol.* **15,** S7–S16.
18. Anisowicz, A., Sotiropoulou, G., Stenman, G., Mok, S. C., and Sager, R. (1996) *Mol. Med.* **2,** 624–636.
19. Little, S. P., Dixon, E. P., Norris, F., Buckley, W., Becker, G. W., Johnson, M., Dobbins, J. R., Wyrick, T., Miller, J. R., MacKellar, W., Hepburn, D., Corvalan, J., McClure, D., Liu, X., Stephenson, D., Clemens, J., and Johnstone, E. M. (1997) *J. Biol. Chem.* **272,** 25135–25142.
20. Luo, L., Herbrick, J. A., Scherer, S. W., Beatty, B., Squire, J., and Diamandis, E. P. (1998) *Biochem. Biophys. Res. Commun.* **247,** 580–586.
21. Nelson, P. S., Gan, L., Ferguson, C., Moss, P., Gelinas, R., Hood, L., and Wang, K. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 3114–3119.
22. Yamashiro, K., Tsuruoka, N., Kodama, S., Tsujimoto, M., Yamamura, Y., Tanaka, T., Nakazato, H., and Yamaguchi, N. (1997) *Biochim. Biophys. Acta* **1350,** 11–14.
23. Yoshida, S., Taniguchi, M., Hirata, A., and Shiosaka, S. (1998) *Gene* **213,** 9–16.
24. Yousef, G. M., and Diamandis, E. P. (1999) *J. Biol. Chem.* **274,** 37511–37516.
25. Yousef, G. M., Obiezu, C. V., Luo, L. Y., Black, M. H., and Diamandis, E. P. (1999) *Cancer Res.* **59,** 4252–4256.
26. Stephenson, S. A., Verity, K., Ashworth, L. K., and Clements, J. A. (1999) *J. Biol. Chem.* **274,** 23210–23214.
27. Riegman, P. H., Vlietstra, R. J., Suurmeijer, L., Cleutjens, C. B., and Trapman, J. (1992) *Genomics* **14,** 6–11.
28. Nadeau, J. H., Davisson, M. T., Doolittle, D. P., Grant, P., Hillyard, A. L., Kosowsky, M., and Roderick, T. H. (1991) *Mamm. Genome* **1,** S461–S515.
29. Saunders, A. M., and Seldin, M. F. (1990) *Genomics* **8,** 525–535.
30. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed. Cold Spring Harbor Laboratory, Plainview, NY.
31. Osoegawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J. J., and de Jong, P. J. (1998) *Genomics* **52,** 1–8.
32. Birren, B., Green, E. D., Klapholz, S., *et al.* (1998) Genome Analysis: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Plainview, NY.
33. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.* **25,** 3389–3402.
34. Hansson, L., Stromqvist, M., Backman, A., Wallbrandt, P., Carlstein, A., and Egelrud, T. (1994) *J. Biol. Chem.* **269,** 19420–19426.
35. Liu, X. L., Wazer, D. E., Watanabe, K., and Band, V. (1996) *Cancer Res.* **56,** 3371–3379.
36. Yousef, G. M., Scorilas, A., and Diamandis, E. P. (2000) *Genomics* **63,** 88–96.
37. Yoshida, S., Taniguchi, M., Suemoto, T., Oka, T., He, X., and Shiosaka, S. (1998) *Biochim. Biophys. Acta* **1399,** 225–228.
38. Yousef, G. M., Luo, L. Y., Scherer, S. W., Sotiropoulou, G., and Diamandis, E. P. (1999) *Genomics* **62,** 251–259.
39. Yousef, G. M., Luo, L. Y., and Diamandis, E. P. (1999) *Anticancer Res.* **19,** 2843–2852.
40. Morris, B. J., Catanzaro, D. F., Richards, R. I., Mason, A., and Shine, J. (1981) *Clin. Sci.* **61**(Suppl. 7), 351s–353s.
41. Richards, R. I., Catanzaro, D. F., Mason, A. J., Morris, B. J., Baxter, J. D., and Shine, J. (1982) *J. Biol. Chem.* **257,** 2758–2761.
42. van Leeuwen, B. H., Evans, B. A., Tregear, G. W., and Richards, R. I. (1986) *J. Biol. Chem.* **261,** 5529–5535.
43. Evans, B. A., and Richards, R. I. (1985) *EMBO J.* **4,** 133–138.
44. Irwin, D. M., Robertson, K. A., and MacGillivray, R. T. (1988) *J. Mol. Biol.* **200,** 31–45.
45. Fortier, A. H., Nelson, B. J., Grella, D. K., and Holaday, J. W. (1999) *J. Natl. Cancer Inst.* **91,** 1635–1640.
46. Stenman, U. H. (1999) *Clin. Chem.* **45,** 753–754.
47. Goyal, J., Smith, K. M., Cowan, J. M., Wazer, D. E., Lee, S. W., and Band, V. (1998) *Cancer Res.* **58,** 4782–4786.
48. Tanimoto, H., Underwood, L. J., Shigemasa, K., Yan Yan, M. S., Clarke, J., Parmley, T. H., and O'Brien, T. J. (1999) *Cancer* **86,** 2074–2082.
49. Underwood, L. J., Tanimoto, H., Wang, Y., Shigemasa, K., Parmley, T. H., and O'Brien, T. J. (1999) *Cancer Res.* **59,** 4435–4439.