Published in final edited form as:

Gene. 2007 May 1; 392(1-2): 142–150.

Insights into extensive deletions around the XK locus associated with McLeod phenotype and characterization of two novel cases

Jianbin Peng¹, Colvin Redman¹, Xu Wu¹, Xiaoling Song¹, Ruth H. Walker², Connie M. Westhoff³, and Soohie Lee¹

¹ New York Blood Center, New York, NY 10021, USA

² James P. Peters VAMC, Bronx, NY 10468, USA

³ American Red Cross, Philadelphia, PA 19123, USA

Abstract

The McLeod phenotype is derived from various forms of *XK* gene defects that result in the absence of *XK* protein, and is defined hematologically by the absence of Kx antigen, weakening of Kell system antigens, and red cell acanthocytosis. Individuals with the McLeod phenotype usually develop late-onset neuromuscular abnormalities known as the McLeod syndrome (MLS). MLS is an X-linked multi-system disorder caused by absence of *XK* alone, or when the disorder is caused by large deletions, it may be accompanied with Duchenne muscular dystrophy (*DMD*), chronic granulomatous disease (*CYBB*), retinitis pigmentosa (*RPGR*), and ornithine transcarbamylase deficiency (*OTC*).

XK defects derived from a large deletion at the *XK* locus (Xp21.1) have not been characterized at the molecular level. In this study, the deletion breakpoints of two novel cases of McLeod phenotype with extensive deletions are reported. Case 1 has greater than 1.12 million base-pairs (mb) deletion around the *XK* locus with 7 genes affected. Case 2 has greater than 5.65 mb deletion from *TCF7L1* to *DMD* encompassing 20 genes. Phylogenetic analyses demonstrated that *DMD*, *XK* and *CYBB* have close paralogs, some of which may partially substitute for the functions of their counterparts. The loci around *XK* are highly conserved from fish to human; however, the disorders are probably specific to mammals, and may coincide with the translocation of the loci to the X chromosome after the speciation in birds. The non-synonymous to synonymous nucleotide substitution rate ratio ($\omega = dN/dS$) in these genes was examined. *CYBB* and *RPGR* show evidence of positive selection, whereas *DMD*, *XK* and *OTC* are subject to selective constraint.

Keywords

X-linked disorders; Duchenne muscular dystrophy; chronic granulomatous disease; retinitis pigmentosa; ornithine transcarbamylase deficiency; adaptive evolution

Correspondence: Soohie Lee, PhD. The New York Blood Center 310 East 67th Street New York, NY 10021 Telephone: 212-570-3062 Fax: 212-897-0243 E-mail: solee@nybloodcenter.org

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. INTRODUCTION

XK is located at Xp21.1, contains 3 exons and is over 50 kb in size. The gene product, *XK* a 50 kDa protein, is predicted to span cellular membranes 10 times and because of its physical characteristics is presumed to be a putative membrane transport protein with unknown substrate. Absence of *XK* in the McLeod phenotype, results in clinical manifestations known as the McLeod syndrome (MLS) that affects hematology, as well as the central and peripheral nervous systems in middle-aged men. The McLeod phenotype is characterized by very weak expression of Kell erythrocyte antigens, red cell acanthocytosis, and elevated levels of serum creatine phosphokinase (Redman et al 1999). The clinical features of MLS are heterogeneous varying even between siblings, and include limb chorea, seizures, peripheral neuropathy, muscle atrophy, cardiomyopathy and psychopathology. It may be difficult to distinguish MLS from other neuromuscular disorders by clinical features and traditional examination (Starling et al 2005; Wada et al 2003). Determination can be achieved by observing the absence of the red cell Kx antigen, but *XK* genotyping leads to final confirmation of the syndrome.

MLS is an X-linked recessive disorder resulting from various *XK* gene mutations or deletions. *XK* mutations include single nucleotide mutations at or near the splice junctions of introns resulting in aberrant proteins or reduced amount of *XK* protein (Danek et al 2001; Walker et al 2006a). Single nucleotide mutations are also present in the coding region resulting in an amino acid substitution or introducing a premature stop codon that leads to truncated *XK*. In the case of amino acid substitutions, an *XK* variant was shown not to travel to the plasma membrane (Russo et al 2002). *XK* deletions may be small, ranging from a single to a few nucleotides which result in frame shifts, or may be large, resulting in partial deletion or complete deletion of *XK*, which may include also neighboring genes (Danek et al 2001; Walker et al 2006a).

To date the majority of reported cases of MLS are due to mutations or deletions within the *XK* gene (Danek et al 2001; Rampoldi et al 2002). Large deletions around the *XK* locus may cause a combination of MLS with Duchenne muscular dystrophy (DMD) resulting from mutations in *DMD*, chronic granulomatous disease (CGD) resulting from mutations in *CYBB*, X-linked retinitis pigmentosa (XLRP) resulting from mutations in *RPGR*, or possibly ornithine transcarbamylase deficiency (OTC deficiency) resulting from *OTC* mutations (El et al 2000). However, cases where the majority of the five genes are simultaneously deficient are very rare.

In this study we report two unrelated cases of MLS with extensive deletions that are novel with regard to their sizes and deletion positions. Previous studies of McLeod phenotypes have reported large deletions that include *XK* and neighboring genes (*DMD*, *CYBB*, *RPGR*, and *OTC*) (Bertelson et al 1988; Ng et al 2004; Wendel et al 2004), but determinations of the exact breakpoints and systematic analyses of the five genes in the region are not reported.

In this study we conducted bioinformatics based analyses for the following purposes: First, deficiencies of the five genes all cause X-linked disorders in human. Some disorders are also reported in mice (Bingel 2002; Hong et al 2000; Nonaka 1998; Ratnakumari et al 1996), but the evolution of the chromosomal segment comprising these genes from lower to higher species remains unknown. Analysis of gene linkage from fish to human may provide a picture of the evolution of the chromosomal segment. Second, naturally occurring mutations in humans and gene knockout studies in mice confirmed that the functional losses of these genes (but not all simultaneously) are not immediately lethal, indicating that the functions of some of these genes might be compensated or substituted by their paralogs (Lopez-Bigas and Ouzounis 2004; Tinsley et al 1998). Therefore, we conducted phylogenetic analyses of the five genes, and identified the close and distant paralogs (if any exist). Third, understanding the molecular evolution pattern of the five genes in terms of adaptive evolution (positive selection) resulting

from advantageous mutations, or selective constrain (purifying selection or negative selection) resulting from deleterious mutations, might provide insight into the function of these genes, specially XK whose function is not known.

2. MATERIALS AND METHODS

2.1. Preparation of genomic DNA

Blood samples from two McLeod cases were obtained with informed patient consent. Genomic DNA was extracted from the blood sample of case 1, his brother and his mother by the procedure described previously (John et al 1991; Lee et al 1995), and DNA was isolated from the blood samples of case 2 and his mother with the QIAamp Blood Mini Kit (QIAGEN, Hilden, Germany).

2.2 Localization of the proximal deletion breakpoints by PCR

Initial XK genotyping using 3 sets of PCR primers that amplify the 3 exons and splice junctions of XK indicated that both of the McLeod phenotypes are derived from gene deletions. In case 1, only exon 3 of XK amplified and in case 2 no amplification products were obtained, indicating that the entire XK was deleted. In order to determine the extent of the gene deletions, a series of PCR primers from sequences in the neighboring genes spanning approximately 8 million base-pairs (mb) from *OTC* to the *FTHL17* genes were designed and tested. Primers were designed to target unique areas in the human genome sequence, avoiding areas homologous to other regions. Most of the amplifications used Taq polymerase (Promega, Madison WI) but for areas that are hard to amplify due to high GC content, Herculase Hotstart DNA polymerase (Stratagene, La Jolla, CA) was used following the manufacturer's protocol. Normal genomic DNA was used as a positive PCR control. The primer sequences are listed in the supplementary document. Following the identification of the proximal deletion breakpoint from either 3' or 5' side of the deleted region one set of the nested gene specific primers was designed from the proximal breakpoint of the unaffected region in order to amplify the affected region of the DNA segment from an adaptor-ligated genomic DNA pool made from the genomic DNA of the two McLeod phenotype subjects.

2.3 Amplification of the affected gene segment

Following digestion of genomic DNA with the restriction enzymes (*EcoRI*, *DraI*, *PvuII* and *SspI*) in the BD Genome Walker kit (Clontech, Mountain View, CA), pools of adaptor-ligated genomic DNA fragments were constructed following the manufacturer's protocol. The genomic DNA segments around the affected regions for each of the cases were amplified by nested PCR using the gene specific nested primers and nonspecific nested adopter primers provided in the kit. The resulting PCR products were subcloned and the sequence was compared to the normal genomic sequence retrieved from human genome database at GenBank to identify the exact deletion breakpoints. The outer and nested gene specific primer (GSP) sequences and a list of the preceding PCR primer sequences are available in the supplementary document. The deletion breakpoint was confirmed for both cases by amplifying the affected area using the primers designed to include the breakpoint and sequencing the PCR product. Presence of the affected allele in the mothers of both cases was also similarly checked by PCR to detect the inheritance of the affected gene.

2.4 Bioinformatics-based analysis of the disorder-related genes at Xp21

Genomic sequences comprising multiple genes around the XK locus were retrieved from genome databases of human, dog, mouse, chicken, green pufferfish at NCBI (<http://www.ncbi.nlm.nih.gov/blast/>). Protein/cDNA sequences of DMD, XK, CYBB, RPGR and OTC, and their related paralogs (if any exist) were retrieved from NCBI databases using

BLAST search (Altschul et al 1997). The sequences used for the analysis were inspected for errors against available expressed sequence tags (ESTs), mRNAs and genomic contigs. Protein sequences were aligned using MUSCLE (version 3.52) (Edgar 2004), then codon sequence alignments were derived. Phylogenetic analysis was conducted using the Neighbor Joining (NJ) method in Mega (version 3.0) (Kumar et al 2001). Analysis of adaptive evolution of the five genes was carried out using CODEML (Yang et al 2000) in PAML package (version 3.15) (Yang 1997), where non-synonymous nucleotide substitution (missense mutation) to synonymous nucleotide substitution (silent mutation) rate ratio (dN/dS) between homologous protein-coding gene sequences were analyzed. The resulting ratio, dN/dS, <1, =1 and >1 indicated purifying selection, neutral evolution, and positive selection, respectively. The dataset and sequence alignments are provided in the supplementary document.

3. RESULTS AND DISCUSSION

3.1 RBC phenotype

Case 1—The red cells from case 1 were negative when tested with anti-KL. Anti-KL contains both anti-Km and anti-Kx. Red cells from case 1 and his brother reacted weakly with the antibodies to common Kell antigens. Case 1 was K–k+w and his brother was K+w k+w (Marsh et al 1983). Western blot analysis of the red cell membrane proteins, using rabbit antibody to XK C-terminal region peptide (XK 404–444), did not detect XK, truncated XK or Kell/XK complex.

Case 2—Red cells of case 2 typed K–/k–, Kp(a–b–), Ku–, Kx– and were also non-reactive with anti-Km. Expression of Kell antigens in McLeod phenotypes are typically detected as weak reactions, with the exception of inheritance of a *KEL3* (*Kp^a*) allele which can further reduce Kell antigen expression (Daniels et al 1996). The apparent absence of Kell antigens on the red cells of this child was not due to inheritance of *KEL3*, as the genotype was *KEL4*/*KEL4*, i.e. homozygous for *Kp^b*. The apparent absence of Kell antigens requires further analysis of the *KEL* genes and serologic confirmation by testing with other source antibodies and by adsorption and elution studies with the red cells.

3.2 Gene analyses

Case 1—The DNA segment amplified from a pool of adaptor-ligated genomic DNA fragments following digestion of genomic DNA with *PvuII* was subcloned and sequenced to identify the breakpoint (Fig. 1A). The deleted region is 1.12 mb and spans from the second intron of *XK* towards the 5' direction and up to *LOC441488*. Seven protein-coding genes and 3 pseudo-genes are affected in the region (Fig. 1B). The DNA segment around the deletion breakpoint of case 1, his affected brother and his mother were amplified (Fig. 1C) showing that the affected allele was inherited.

Case 2—The DNA segment amplified from a pool of adaptor-ligated genomic DNA fragments following DNA digestion with *DraI* was subcloned and sequenced to identify the exact deletion breakpoint (Fig. 2A). More than 5.65 mb is deleted around the *XK* locus. The deleted region spans from *TCTEIL* to *DMD*. Twenty protein-coding genes and 4 pseudo-genes are affected in this region, including *CYBB*, *XK*, and *DMD* genes (Fig. 2B). The DNA segment around the deletion breakpoint of case 2 and his mother were amplified (Fig. 2C) showing that the affected allele was inherited.

3.3 Clinical descriptions

Case 1—This patient was incidentally noted to have acanthocytosis and McLeod Kell phenotype at age 30 (Marsh et al 1983). He first noted problems with walking due to leg weakness aged 50, and had a first seizure five years later. On examination at age 56 he reported

occasional auras, tongue biting, bruxism, leg restlessness, and memory problems. He had very mild leg chorea. Speech was mildly dysarthric. Power and tone were diminished throughout. Deep tendon reflexes were absent throughout. He pushed up to stand due to proximal muscle weakness; gait was shambling with bilateral footdrop. Echocardiography showed minimal aortic and mitral valve dysfunction, abnormal left ventricular filling pattern but normal size and function. He had been diagnosed with scleroderma, but did not show any increased tendency to infection. His brother suffered from cardiac disease, mild chorea and peripheral neuropathy (Walker et al 2006b), and did not have any signs or symptoms of autoimmune disease, nor tendency to infection.

Case 2—This patient was diagnosed with CGD at age 4 months. He presented with mild seizure-like activity requiring hospitalization, a skin infection in the groin, and impetigo. Laboratory tests showed a mild elevation in transaminases and the patient had hepatosplenomegaly. NBT (Nitroblue tetrazolium) testing was consistent with CGD. He was treated with immunoglobulin, interferon, and Bactrim (trimethoprim-sulfamethoxazole), and has recently undergone bone marrow transplantation at the age of 23 months. He did not have any other identifiable abnormalities other than those listed above. The symptoms of DMD and MLS both typically develop at an older age (childhood and mid-adulthood, respectively) than this infant, thus the absence of symptoms related to the *XK* and *DMD* gene defects is not surprising.

3.4 Gene linkages around the *XK* locus

The chromosomal region around the *XK* locus from vertebrate fish to human is highly conserved (Fig. 3A). *LANCL3*, *XK*, *TCTEIL*, *SRPX* and *OTC* genes are close neighbors throughout this range of species. In green pufferfish, the assembly of chromosomes has not been completed at the present time. The above five genes are located at the same genomic contig on chromosome 2. *CYBB*, *TCTEIL**, *RPGR* resides on another contig on chromosome 3, *DMD* is also on chromosome 3 but on a different contig, and *SYTL5* is on chromosome 1. Two *TCTEIL* copies labeled as *TCTEIL* and *TCTEIL** are found in green pufferfish, and both share 71% identity at the protein level. *TCTEIL* (next to *XK*) has greater identity with *TCTEIL* in chicken, mouse, dog and human than does the *TCTEIL** (next to *CYBB*). In chicken, *CYBB*, *RPGR*, *DMD* and *SYTL5* are in the same loci as *XK* and the change in gene order in the region are minor from chicken to human. The largest position change relative to the *XK* locus is *DMD* in mouse, which is located approximately 71 mb downstream of *XK*, while in human and dog it is located approximately 6.4 mb and 5.0 mb upstream of *XK*, respectively.

Large deletions in the McLeod cases may extend from *XK* to *DMD*, as in case 2 reported here, and can also occur on the other side of *XK* to *OTC*. The 6.4 mb region from *XK* to *DMD* harbors more genes than the 0.7 mb region from *XK* to *OTC*. However, it is interesting to note that four of the five disease related genes, *XK*, *CYBB*, *RPGR* and *OTC*, reside in the short 0.7 mb segment. The incidence of mutations and/or deletions around the *XK* locus in species other than human remains virtually unknown but is probably different between mammals and lower species. This segment containing *XK* is located on an autosome from fish to chicken, whereas in mammals (mouse, rat, dog, chimp and human) this segment is on the X chromosome (Fig. 3A). The genetic mechanism by which a deficient gene is passed from parents to offspring is different in autosomal and sex chromosomes, thus leading to different phenotypic consequences between males and females in humans, and presumably in other mammals. The genomic segments that are deleted in the McLeod cases are located on the short arm of X chromosome at Xp21 where *XK* resides. However, the long arm of the X chromosome around Xq22 contains *XPLAC*, which is the paralog of *XK*. This segment also harbors genes that are homologous to genes around the *XK* locus (Fig. 3B). *SRPX* and *SRPX2*, *SYTL5* and *SYTL4*, *CYBB* and *NOX1*, *XK* and *XPLAC*, *DMD* and *DRP2* are close pairs of paralogs. *DRP2* has

sequence similarity to the carboxy-terminal of *DMD*. The other four pairs are complete gene duplications.

As large deletions in *MLS* are often extended to other neighboring genes, the similar patterns of gene linkages in higher animals like chimp and dog, suggest that the large deletion in *XK* locus might not be human specific.

3.5 Phylograms and paralogs functionally related to the five genes

In order to gain insights into the functional relationships of the homologs within each of the five disorder related genes, we conducted phylogenetic analyses (related phylograms are provided in the supplementary document). Gene duplications can be found within families of *DMD*, *XK* and *CYBB*, whereas *RPGR* and *OTC* do not have paralogs indicating that there was no gene duplication or that the paralogs were lost after gene duplication. Orthologs may retain the same function in the course of evolution in different species, whereas paralogs may develop new functions that are related to the original functions of the primary gene in a genome. The consequences of gene duplication in a genome may lead to pseudo-, sub- and neo-functionalization, as well as to functional conservation by which the duplicate may provide extra amounts of proteins when the product is in high demand (Zhang 2003a). As a result, proteins with similar paralogs would be less likely to be involved in disease as they could rescue the mutant phenotype (Lopez-Bigas and Ouzounis 2004). Human *DMD* and *UTRN* share about 46% identity in protein sequences. *UTRN* is able to substitute some of the functions of *DMD* in *DMD*-deficient mice, and there is evidence that *UTRN* and *DMD* can perform similar functions in the sarcolemma (Blake et al 2002; Tinsley et al 1998). A phylogenetic tree demonstrates that *UTRN* is a close paralog of *DMD*, found as early as in *Ciona intestinalis* and *Strongylocentrotus purpuratus*. Although gene *DRP2* (Xq22, 954aa, NM_001939) resembles C-terminal isoforms of *DMD* and *UTRN*, and is possibly duplicated from *DMD*, it lacks the majority of *DMD* coding region and probably cannot substitute for the functions of *DMD*.

XPLAC and *XTES* are two paralogs of the *XK* gene. At the protein sequence level, human *XK* shares 38% identity with *XPLAC*, and 32% with *XTES*. Phylogenetic analysis indicates that *XK* is the primary copy; that *XPLAC* is duplicated from *XK*; and that *XTES* is later duplicated from *XPLAC*, around the time that primates originated (Calenda et al 2006). *XTES* is exclusively expressed in human testis, but *XK* and *XPLAC* have overlapping expression in some tissues (Calenda et al 2006). *CYBB* has four paralogs, including *NOX1*, *NOX3*, *NOX4* and *NOX5*. Human *CYBB* shares 58% protein identity with *NOX1*, 56% with *NOX3*, 36% with *NOX4*, and 20% with *NOX5*. However, functional substitution of *NOX1* or other paralogs for *CYBB* has not been established.

RPGR and *OTC* have no identifiable paralogs within the same organisms from fish to human. *RPGR* can be found in vertebrate fish; and *OTC* in *Strongylocentrotus purpuratus*. *RPGR*-deficiency appears to have an early age of onset (Hong et al 2000); but *OTC*-deficiency is enigmatic as the age of onset is often variable (van Diggelen et al 1996). It is not known if other non-sequence related genes to *RPGR* and *OTC* may serve similar functions.

On red blood cells, Kell and *XK* are covalently linked by a single disulfide bond. Thus, they may be considered to be subunits of a single protein having a complementary function (Russo et al 1998). *KEL*, in contrast to *XK*, is only found in birds and higher species after duplication from its ancestor gene *ECE1/ECE2* (unpublished data); *XK* evolved as early as in vertebrate fishes. This suggests that *XK* as a complex underwent a functional shift when it entered warm-blood species.

3.6 Selective pressure on DMD, XK, CYBB, RPGR and OTC genes

Genes on the X chromosome are often of particular interest due to their roles in sexual dimorphism (Pask and Graves 1999) and adaptive evolution is often an important topic regarding these genes. The five disorder-related genes (*XK*, *DMD*, *CYBB*, *RPGR* and *OTC*) were examined for selective patterns in molecular evolution based on the nucleotide mutations present in coding regions. Four site-based codon models (Wong et al 2004; Yang 1997; Yang et al 2000) were used for the analysis: Nearly neutral model (M1a), positive selection model (M2a), beta model (M7), and beta plus ω model (M8, and null hypothesis M8a where beta & $\omega=1$) (Table 1). Under these site-based codon models, *DMD*, *XK* and *OTC* were found to be under purifying selection, since $dN/dS (= \omega)$ values over most codon sites were far less than 1, and no positively selected sites were identified. *CYBB* and *RPGR* were found to be under positive selection. For *CYBB*, the M2a model identified 1.6% of codon sites at positive selection with $\omega = 3.6$ and $p\text{-value}=0.0074$ against M1a. The M8 model identified 3.6% of codon sites at positive selection with $\omega=2.2$ and $p\text{-value}=0.0001$ against M8a (Table 1). Similarly for *RPGR*, 6.1% of sites at $\omega=3.8$ under M2a model; and 17.8% of sites at $\omega=2.0$ under M8 model; and the $p\text{-values}$ of M2a vs M1a, and M8 vs M7, and M8 vs M8a were <0.0001 .

In order to identify potential episodes of positive selection during evolutionary stages for the *XK* gene, we further analyzed the same dataset of *XK* genes by applying a branch-site model (Yang 1997; Yang et al 2000). The data revealed that from chicken to human: $p\text{-value}=1.0000$ for the branch leading to mouse/rat, and the branch to dog/cow, respectively; $p\text{-value}=0.8187$ for the branch to chimp/human. These results indicate that *XK* did not experience positive selection as it evolved from chicken to human.

CYBB and *RPGR* were subject to positive selection, not surprisingly since *CYBB* is associated with the immune system, and *RPGR* is related to the visual sensory system and both have undergone rapid development during mammalian evolution. Although *XK* is part of the Kx blood group system and *XK*-deficiency is associated with neuromuscular abnormalities, its selection pattern is different from that of other blood group genes such as *ABO* and *RHCE*, and different from *ASPM* (abnormal spindle-like microcephaly associated) which is involved in brain size. The latter three genes were all verified to be under positive selection (Huang and Peng 2005; Peng and Huang 2006; Saitou and Yamamoto 1997; Zhang 2003b). Because of the resemblance of *MLS* to Huntington's disease and chorea-acanthocytosis, caused by mutations in *HTT* (*HD*) and *VPS13A* (*CHAC*), respectively, it was suggested that the three corresponding gene products (*XK*, huntingtin, and chorein) may belong to a common functional pathway, since their dysfunctions all lead to degeneration of the basal ganglia (Danek et al 2001). Beside *XK*, we found that *VPS13A* and *HTT* genes are also subject to purifying selection (Table 1) demonstrating that the three genes (*XK*, *VPS13A* and *HTT*) with similar phenotypes are all under selective constraint.

The close paralogs of *XK*, *CYBB* and *DMD* were also examined for molecular selection patterns. Interestingly, *XPLAC*, a paralog of *XK*, is under positive selection and *NOX1*, a paralog of *CYBB* is under purifying selection; For the two pairs of close paralogs, *XK/XPLAC* and *CYBB/NOX1*, one paralog is in purifying selection with its counterpart in positive selection and vice versa. The selection pattern of *UTRN*, a paralog of *DMD* is inconclusive (table 1). Positive selection acts on a gene only when the gene function is altered and the organisms' fitness to selection pressure is increased (Zhang 2003b), thus the result here provides some insight into the functions of these genes. *XK* is the primary copy and is under purifying selection, whereas the duplicate, *XPLAC*, is subject to positive selection. *XK* is widely expressed in many different tissues while *XPLAC* is not as widely expressed as *XK*, however they do overlap in some tissues and share 38% amino acid sequence identities, and similar secondary structures. These facts suggest that during evolution *XK* maintained certain basic functions that are widely required, while *XPLAC* probably developed new functions but still preserved some vital

functions that overlap with XK. In comparison with XK, XPLAC is expressed in a larger amount in thymus (unpublished data), which plays an important role in the development of the immune system in early life, with its cells forming a part of the body's normal immune system. We speculate that the positive selection of XPLAC is related to its possible functions in the mammalian immune system. Nine positively selected sites (30S, 204S, 398L, 440Q, 441H, 442P, 443R, 444T, 446V; positions referred to NP_997724) were found in mammalian XPLAC by Bayes Empirical Bayes analysis in M8 model. Five of these nine sites are presented in internal C-terminal domain implying that this domain may be responsible for a unique function of XPLAC which is different from XK function.

The average value of the non-synonymous substitution rate in *XK* is much lower than that of *XPLAC* (Fig. 4); whereas in chimp and human, a new homolog *XTES* was duplicated from *XPLAC*, and shows higher non-synonymous substitution rate than *XPLAC*. Comparing coding sequences between human and chimp demonstrates that *XTES* has the highest frequency of nucleotide substitutions, indicating that *XTES*, as a newcomer is probably developing new functions based on those exhibited by *XPLAC*.

It is difficult to establish suitable animal models for mimicking extensive deletions corresponding to multiple gene loss. Mouse models of intragenic defect for *DMD*, *CYBB*, *RPGR* and *OTC* have been reported. *DMD*-deficient mice have elevated plasma levels of muscle creatine kinase and pyruvate kinase and exhibit histological lesions characteristic of muscular dystrophy (Bulfield et al 1984); but the pathophysiology of the disease in *DMD*-mutant mice does not mimic entirely the human condition resulting from defects in *DMD* (Chambers et al 2001). Targeted disruption of single genes has proved useful and has provided pertinent information on their physiological functions. For example, *CYBB*-knockout mice lacking the b subunit of NADPH oxidase are susceptible to experimental infection with *Aspergillus fumigatus* (Bingel 2002), and *RPGR*-mutant mice exhibit ectopic localization of cone opsins in the cell body and synapses of cone photoreceptors and have a reduced level of rhodopsin in rod photoreceptors (Hong et al 2000). *OTC*-deficiency mice exhibit significantly lower levels of arginine in the brain (Ratnakumari et al 1996). These models may enrich our understanding of the functions of related genes. However, integrated functions from multiple genes probably are not only a simple concatenation of the functions of individual genes. Multiple gene loss may add much complexity to the system, which may become more cryptic, while functions of some defect-genes may be substituted by corresponding paralogs under certain conditions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors thank Dr. Marion Reid of the Immunohematology Laboratory, New York Blood Center, for providing frozen red cells of case 1, and the staff of the Nucleic Acid Analysis Laboratory, New York Blood Center for DNA sequencing. The study was supported in part by an NIH Specialized Center of Research (SCOR) grant in Transfusion Biology and Medicine, (HL54459), and by an NIH grant (RO1 HL075716).

References

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–3402. [PubMed: 9254694]
- Bertelson CJ, Pogo AO, Chaudhuri A, Marsh WL, Redman CM, Banerjee D, Symmans WA, Simon T, Frey D, Kunkel LM. Localization of the McLeod locus (XK) within Xp21 by deletion analysis. *Am J Hum Genet* 1988;42:703–711. [PubMed: 3358422]

- Bingel SA. Pathology of a mouse model of x-linked chronic granulomatous disease. *Contemp Top Lab Anim Sci* 2002;41:33–38. [PubMed: 12213046]
- Blake DJ, Weir A, Newey SE, Davies KE. Function and genetics of dystrophin and dystrophin-related proteins in muscle. *Physiol Rev* 2002;82:291–329. [PubMed: 11917091]
- Bulfield G, Siller WG, Wight PA, Moore KJ. X chromosome-linked muscular dystrophy (mdx) in the mouse. *Proc Natl Acad Sci U S A* 1984;81:1189–1192. [PubMed: 6583703]
- Calenda G, Peng J, Redman CM, Sha Q, Wu X, Lee S. Identification of two new members, XPLAC and XTES, of the XK family. *Gene* 2006;370:6–16. [PubMed: 16431037]
- Chambers SP, Dodd A, Overall R, Sirey T, Lam LT, Morris GE, Love DR. Dystrophin in adult zebrafish muscle. *Biochem Biophys Res Commun* 2001;286:478–483. [PubMed: 11511083]
- Danek A, et al. McLeod neuroacanthocytosis: genotype and phenotype. *Ann Neurol* 2001;50:755–764. [PubMed: 11761473]
- Daniels GL, Weinauer F, Stone C, Ho M, Green CA, Jahn-Jochem H, Offner R, Monaco AP. A combination of the effects of rare genotypes at the *XK* and *KEL* blood group loci results in absence of Kell system antigens from the red blood cells. *Blood* 1996;88:4045–4050. [PubMed: 8916972]
- Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;32:1792–1797. [PubMed: 15034147]
- El NW, Colin Y, Collec E, Gane P, Cartron JP, Kim CL. Analysis of deletions in three McLeod patients: exclusion of the XS locus from the Xp21.1-Xp21.2 region. *Eur J Immunogenet* 2000;27:29–33. [PubMed: 10651848]
- Hong DH, Pawlyk BS, Shang J, Sandberg MA, Berson EL, Li T. A retinitis pigmentosa GTPase regulator (RPGR)-deficient mouse model for X-linked retinitis pigmentosa (RP3). *Proc Natl Acad Sci U S A* 2000;97:3649–3654. [PubMed: 10725384]
- Huang CH, Peng J. Evolutionary conservation and diversification of Rh family genes and proteins. *Proc Natl Acad Sci U S A* 2005;102:15512–15517. [PubMed: 16227429]
- John SW, Weitzner G, Rozen R, Sriver CR. A rapid procedure for extracting genomic DNA from leukocytes. *Nucleic Acids Res* 1991;19:408. [PubMed: 2014181]
- Kumar S, Tamura K, Jakobsen IB, Nei M. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 2001;17:1244–1245. [PubMed: 11751241]
- Lee S, Wu X, Reid M, Zelinski T, Redman C. Molecular basis of the Kell (K1) phenotype. *Blood* 1995;85:912–916. [PubMed: 7849312]
- Lopez-Bigas N, Ouzounis CA. Genome-wide identification of genes likely to be involved in human genetic disease. *Nucleic Acids Res* 2004;32:3108–3114. [PubMed: 15181176]
- Marsh WL, Schnipper EF, Johnson CL, Mueller KA, Schwartz SA. An individual with McLeod syndrome and the Kell blood group antigen K(K1). *Transfusion* 1983;23:336–338. [PubMed: 6879675]
- Ng D, et al. Oculofaciocardiodental and Lenz microphthalmia syndromes result from distinct classes of mutations in BCOR. *Nat Genet* 2004;36:411–416. [PubMed: 15004558]
- Nonaka I. Animal models of muscular dystrophies. *Lab Anim Sci* 1998;48:8–17. [PubMed: 9517883]
- Pask A, Graves JA. Sex chromosomes and sex-determining genes: insights from marsupials and monotremes. *Cell Mol Life Sci* 1999;55:864–875. [PubMed: 10412369]
- Peng J, Huang CH. Rh proteins vs Amt proteins: an organismal and phylogenetic perspective on CO2 and NH3 gas channels. *Transfus Clin Biol* 2006;13:85–94. [PubMed: 16564193]
- Rampoldi L, Danek A, Monaco AP. Clinical features and molecular bases of neuroacanthocytosis. *J Mol Med* 2002;80:475–491. [PubMed: 12185448]
- Ratnakumari L, Qureshi IA, Butterworth RF, Marescau B, De Deyn PP. Arginine-related guanidino compounds and nitric oxide synthase in the brain of ornithine transcarbamylase deficient spf mutant mouse: effect of metabolic arginine deficiency. *Neurosci Lett* 1996;215:153–156. [PubMed: 8899736]
- Redman CM, Russo D, Lee S. Kell, Kx and the McLeod syndrome. *Baillieres Best Pract Res Clin Haematol* 1999;12:621–635. [PubMed: 10895256]
- Russo D, Redman C, Lee S. Association of XK and Kell blood group proteins. *J Biol Chem* 1998;273:13950–13956. [PubMed: 9593744]

- Russo DC, Lee S, Reid ME, Redman CM. Point mutations causing the McLeod phenotype. *Transfusion* 2002;42:287–293. [PubMed: 11961232]
- Saitou N, Yamamoto F. Evolution of primate ABO blood group genes and their homologous genes. *Mol Biol Evol* 1997;14:399–411. [PubMed: 9100370]
- Starling A, Schlesinger D, Kok F, Passos-Bueno MR, Vainzof M, Zatz M. A family with McLeod syndrome and calpainopathy with clinically overlapping diseases. *Neurology* 2005;65:1832–1833. [PubMed: 16344536]
- Tinsley J, Deconinck N, Fisher R, Kahn D, Phelps S, Gillis JM, Davies K. Expression of full-length utrophin prevents muscular dystrophy in mdx mice. *Nat Med* 1998;4:1441–1444. [PubMed: 9846586]
- van Diggelen OP, Zaremba J, He W, Keulemans JL, Boer AM, Reuser AJ, Ausems MG, Smeitink JA, Kowalczyk J, Pronicka E, Rokicki D, Tarnowska-Dziduszko E, Kneppers AL, Bakker E. Asymptomatic and late-onset ornithine transcarbamylase (OTC) deficiency in males of a five-generation family, caused by an A208T mutation. *Clin Genet* 1996;50:310–316. [PubMed: 9007316]
- Wada M, Kimura M, Daimon M, Kurita K, Kato T, Johmura Y, Johkura K, Kuroiwa Y, Sobue G. An unusual phenotype of McLeod syndrome with late onset axonal neuropathy. *J Neurol Neurosurg Psychiatry* 2003;74:1697–1698. [PubMed: 14638894]
- Walker RH, Danek A, Uttner I, Offner R, Reid M, Lee S. McLeod phenotype without the McLeod syndrome. *Transfusion*. 2006aRef Type: In Press
- Walker RH, Jung HH, Tison F, Lee S, Danek A. Phenotypic variation among brothers with the McLeod neuroacanthocytosis syndrome. *Mov Disord*. 2006bRef Type: In Press
- Wendel S, Fontão-Wendel R, Levi JE, Aravechia MG, Bordokan RES, Russo D, Haddad MS. A McLeod phenotype detected by random screening for K:-4 [Kp(b-)] blood donors in Brazil. *Transfusion* 2004;44:1579–1587. [PubMed: 15504163]
- Wong WS, Yang Z, Goldman N, Nielsen R. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics* 2004;168:1041–1051. [PubMed: 15514074]
- Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 1997;13:555–556. [PubMed: 9367129]
- Yang Z, Nielsen R, Goldman N, Pedersen AM. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 2000;155:431–449. [PubMed: 10790415]
- Zhang J. Evolution by gene duplication: an update. *Trends in Ecology and Evolution* 2003a;18:292–298.
- Zhang J. Evolution of the human ASPM gene, a major determinant of brain size. *Genetics* 2003b;165:2063–2070. [PubMed: 14704186]

Abbreviations

MLS	McLeod syndrome
DMD	Duchenne muscular dystrophy
CYBB	chronic granulomatous disease
RPGR	retinitis pigmentosa
OTC	ornithine transcarbamylase deficiency
CGD	chronic granulomatous disease

XLRP	X-linked retinitis pigmentosa
dN/dS	synonymous nucleotide substitution rate ratio
NJ	Neighbor Joining (NJ)
PCR	Polymerase chain reaction
BLAST	Basic Local Alignment Search Tool

Figure 1.

Characterization of the deletion at XK gene locus of case 1. (A) The breakpoint and the neighboring sequence for case 1. The blue sequence belongs to an inter-genetic region that is 405 kb away from *LOC158730*; and the red belongs to the intron 3 region in XK. The bold nucleotides demonstrate the breakpoint. (B) Schematic representation of the XK gene locus with 1.12 mb large deletion. The deleted region is denoted in yellow boxes. The expressed genes are in black arrows; the pseudo-genes are labeled in gray arrows. (C) PCR product amplified across the breakpoint. Lane 1 is a DNA ladder; lane 2 genomic DNA from a normal individual; lane 3 is McLeod case 1; lane 4 his affected brother and lane 5 his mother.

Figure 2.

Characterization of the deletion at *XK* gene locus for case 2. (A) The breakpoint and the neighboring sequence for case 2. The blue sequence is in *DMD*, whereas the red belongs to *TCTEIL*. The bold nucleotides demonstrate the breakpoint. (B) Schematic representation of *XK* gene locus with 5.65 mb large deletion. The deleted region is denoted in yellow boxes. The expressed genes are in black arrows; the pseudo-genes are labeled in gray arrows. (C) PCR amplification of the region with the breakpoint. Lane 1 is a DNA ladder, lane 2 is genomic DNA from a normal individual; lane 3 McLeod case 2 and lane 4 his mother.

Figure 3.

XK gene linkage across species, and paralogs around *XK* and *XPLAC* loci in human. (A) Linkage and relative orientation of genes around *XK*. *XK* gene and its homologs are only present as early as in vertebrate fishes. The five disorder-related genes in the study are marked in red; other neighboring genes shown from fish to human are marked in blue. GenBank accession numbers associated to protein sequences are labeled next to the gene name. Genes that are not conserved from fish to human are indicated by dots, where each dot represents one gene and three dots denote three or more genes. In green pufferfish, assembly of chromosomes is not completed at the present time, and therefore, genomic contigs are shown. (B) The paralogs between the short and long arm in the neighboring regions around *XK* and *XPLAC* loci on the human X chromosome. Five genes (*DMD*, *XK*, *CYBB*, *SYTL5* and *SRPX*) have homologs on

the long arm of the same chromosome. The number of amino acids in each protein is labeled under each gene. The thin arrows indicate the gene's orientation relative to the centromere. Identities of protein sequences between paralogs are denoted in percentages.

Figure 4.

Chart of averaged non-synonymous substitution rate (dN) against species orders in *XK* gene family. The rate was computed by comparing the codon sequence of each gene with the common ancestor of *XK*, *XPLAC* and *XTES*. The ancestor sequence was inferred by CODEML (Yang et al 2000). Numbers in the chart represent IDs of taxa are provided in the supplementary document. Vertical coordinate indicates dN value; and horizontal coordinate denotes the order of crown species in million years (Myr) (Huang and Peng 2005). *Xenopus tropicalis* *XPLAC* (#17, with an arrow) is an outlier but no sequence error was found. Red, *XK*; blue, *XPLAC*; brown, *XTES*. Dashed trend lines were drawn for comparison.

Table 1

Selections of the related genes based on site model^{a,b}

Gene ^c	2 $\Delta\ell^d$ from (M2a vs M1a) and statistic test	2 $\Delta\ell$ from (M8 vs M7) and statistic test	2 $\Delta\ell$ from (M8 vs M8a) and statistic test	Conclusion ($\alpha=0.05$)
<i>DMD</i>	2(-30240.6+30240.6) =0.0, df=2, p=1.0000	2(-30217.5+30218.6) =2.2, df=2, p=0.3329	2(-30217.5+30217.9) =0.8, df=1, p=0.1856	Purifying selection
<i>XK</i>	2(-4649.2+4649.2) =0.0 df=2, p=1.0000	2(-4640.3+4641.7) =2.8 df=2, p=0.2466	2(-4640.3+4640.5) =0.4 df=1, p=0.2636	Purifying selection
<i>CYBB</i>	2(-6617.2+6622.1) =9.8, df=2, p=0.0074 $\omega=3.6$ at 1.6% sites	2(-6605.2+6618.6) =26.8, df=2, p<0.0001 $\omega=2.2$ at 3.6% sites	2(-6605.2+6612.2) =14.0 df=1, p=0.0001 $\omega=2.2$ at 3.6% sites	Positive selection
<i>RPGR</i>	2(-12687.7+12698.5) =21.6, df=2, p<0.0001 $\omega=3.8$ at 6.1% sites	2(-12679.6+12698.4) =37.6, df=2, p<0.0001 $\omega=2.0$ at 17.8% sites	2(-12679.6+12690.3) =21.4, df=1, p<0.0001 $\omega=2.0$ at 17.8% sites	Positive selection
<i>OTC</i>	2(-3423.2+3423.2) =0.0 df=2, p=1.0000	2(-3418.1+3418.1) =0.0 df=2, p=1.0000	2(-3418.1+3418.1) =0.0 df=1, p=1.0000	Purifying selection
<i>CHAC</i>	2(-26047.5+26047.5) =0.0, df=1, p=1.0000	2(-26002.0+26002.0) =0.0, df=1, p=1.0000	2(-6002.0+26002.0) =0.0 df=1 p=1.0000	Purifying selection
<i>HD</i>	2(-30579.6+30579.6) =0.0, df=2, p=1.0000	2(-30506.4+30506.7) =0.6 df=2, p=0.7408	2(-30506.4+30506.4) =0.0, df=1, p=1.0000	Purifying selection
<i>XPLAC</i>	2(-3825.2+3828.6) =6.8 df=2, p=0.0334 $\omega=10.5$ at 0.5% sites	2(-3822.2+3828.8) =13.2 df=2, p=0.0014 $\omega=5.4$ at 1.2%	2(-3822.2+3826.5) =8.6 df=1, p=0.0017 $\omega=5.4$ at 1.2%	Positive selection
<i>NOXI</i>	2(-5434.6+5434.6) =0.0 df=2, p=1.0000	2(-5429.1+5429.7) =1.2 df=2, p=0.5488	2(-5429.1+5429.3) =0.4 df=1, p=0.2635	Purifying selection
<i>UTRN</i>	2(-22685.9+22692.9) =14.0, df=2, p=0.0009 $\omega=602.7$ at 0.1% sites	2(-22663.9+22669.1) =10.4, df=2, p=0.0055 $\omega=1.32$ at 3.5% sites	2(-22663.9+22664.3) =0.8, df=1, p=0.371	inconclusive

^a Genes included in the computation are limited to warm-blood vertebrates (from chicken to human) for balanced divergence; but *XPLAC* is from mouse to human because of sequence availability.

^b All datasets and phylograms used in the computation are provided in the supplementary document.

^c *DMD*, *XK*, *CYBB*, *RPGR* and *OTC* are disorder-related genes among the large genomic segment in this study. *CHAC* and *HD* were suggested that their proteins may have the common pathway with *XK* proteins. *XPLAC*, *NOXI* and *UTRN* are close paralogs of *XK*, *CYBB* and *DMD*, respectively.

^d $2\Delta\ell = 2(\ell_1 - \ell_0)$ where ℓ represents the log likelihood values under the model.