HUMAN HERPESVIRUS 8 IN UGANDA: SEROPREVALENCE IN BLOOD DONORS, GENOME VARIABILITY AND EVOLUTION

A thesis

Presented to the Faculty of Biomedical and Life Sciences

at the University of Glasgow

for the Degree of Doctor of Philosophy

by

Dorothy Nalwanga Kakoola

June 2001

Division of Virology

Church Street

Glasgow G11 5JR

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

SUMMARY

Human herpesvirus 8 (HHV-8), also called Kaposi's sarcoma-associated herpesvirus (KSHV), is associated with all forms of Kaposi's sarcoma (KS), and is considered to be the infectious cause of this disease. Classical KS, which affects mainly elderly men in the Mediterranean region, usually presents as benign paranodular skin lesions, but endemic KS of human immunodeficiency virus- (HIV-) negative individuals in equatorial (East and Central) Africa and KS associated with HIV infection or post-transplantation immunosuppression can be clinically aggressive, affecting internal organs. HHV-8 is also associated with primary effusion lymphomas (PEL) and multicentric Castleman's disease.

KS is a prominent disease in Uganda accounting for approximately 50% of reported cancer cases. Epidemic (HIV-associated) and endemic KS occur in Uganda, each afflicting both adults and children.

The highest seroprevalence of HHV-8, with estimates ranging from 11-77%, has been reported for sub-Saharan Africa, followed by the Mediterranean countries (2-28%) and then Northern Europe, Southeast Asia and the Caribbean countries (2-4%). Estimates of seroprevalence in the USA range from 0-20%. The mode by which HHV-8 is transmitted is not known, but transmission most probably occurs via sexual means in countries where KS is non-endemic, and via non-sexual means in countries where KS is endemic, such as Uganda and Zambia.

The HHV-8 genome consists of a unique region of approximately 140.5 kbp flanked by multiple 801 bp terminal repeats. Two almost complete genome sequences are available, one from an AIDS PEL cell line, BC-1 and one from an AIDS KS lesion. The great majority of the genome is highly conserved, but genes at both ends of the unique region exhibit striking variation. The K1 gene,

at the left end of the genome, encodes a highly variable type 1 membrane protein. Five HHV-8 K1 subtypes, generally called A, B, C, D and E have been identified. Subtype B appears to predominate in Africa, together with a variant (A5) of the A subtype, while subtypes A and C predominate elsewhere in the world. The K15 gene, at the right end of the genome, occurs as two highly diverged alleles, P (predominant) and M (minor), showing only 30% amino acid sequence identity. The P allele is more frequent among HHV-8 genomes, and the M allele is rarer. Evidence for recombination has been observed in 20-30% of HHV-8 strains with almost half of the African strains displaying mosaic genomes.

Although KS is a prominent disease in Uganda, little is known about the prevalence of HHV-8 in the Ugandan population and the characteristics of HHV-8 strains present in this country. The aims of this project were to determine the prevalence of HHV-8 in Ugandan blood donors and to characterise the genomes of HHV-8 strains in Ugandan KS patients. During the course of this work, one report on HHV-8 seroprevalence in Ugandan blood donors and three on K1 subtypes in samples from various parts of the world, including some from Uganda, were published.

Using a combination of two serological tests, ELISA (against lytic (ORF65) and latent (ORF73/LANA) antigens) and LANA IF test, antibodies were detected in a high percentage (74%) of 114 HIV-negative blood donors, confirming previous results. Attempts to detect HHV-8 DNA by PCR in peripheral white blood cells of the blood donors were largely unsuccessful.

PCR, sequence and phylogenetic analysis of the K1 gene in tumour samples from 17 KS patients indicated that the majority (11) were infected by the B subtype of HHV-8. Five patients were infected by the A5 subtype. A variant of the C

subtype, showing characteristics of both A and C subtypes in addition to unique characteristics, was detected in a single patient. PCR and sequence analysis of other genome loci implied that at least five of nine (55%) HHV-8 strains are recombinants between subtypes. Three of the five strains with a K1 A5 gene were analysed and all were found to be subtype B in the rest of the genome. The single K1 subtype C strain was also a recombinant between subtypes C and B.

PCR analysis for the K15 gene in tumour samples from 30 KS patients confirmed previous results that the P allele predominates. For the first time, a single strain bearing the M allele was identified in East African samples. Sequence analysis of the entire K15 gene in selected samples indicated that the P and M alleles occur in two distinct forms. Divergence between the M allele found in the Ugandan strain and that in the previously sequenced BC-1 strain is at least as great as that between representatives of the P allele. This indicates that introduction of the M allele into extant HHV-8 subtypes did not occur by a single, relatively recent recombination event as was concluded from a previous study in which very limited variation in the M allele was reported.

The study increases our understanding of the characteristics of HHV-8 strains involved in KS in Uganda and provides new insights into the evolution of HHV-8.

ACKNOWLEDGEMENTS

First and foremost, I thank God for His mercy and grace to me.

I thank Dr. E. Katongole-Mbidde and Dr. P. Kataaha for allowing me access to patients and blood donors, respectively. Without this access the study might never have happened. I thank Nursing sister N. Byabazaire (very sadly now deceased) for her excellent organisational skills and zeal in collecting patients' samples, and Dr. G. Kyeyune and staff at the blood bank for assistance in collecting samples from the blood donors. I am grateful to Prof. D.J. McGeoch for his continuous support and interest in my work throughout the course of these studies. I appreciate the several hours he spent teaching me how to run computer programs for phylogenetic analysis as well the many stimulating discussions we had. I thank him for his words of encouragement and advice during the difficult stage of writing up. He warned me (and in the process encouraged me) that writing up is supposed to be painful. Indeed, it is.

I greatly thank my supervisor, Dr. A.J. Davison, for his excellent supervisory skills. He provided a wonderful and comfortable atmosphere to work in. I thank him for discussions, advice and thorough and critical reading of this manuscript. I also thank him for his help and understanding in times of need. I thank my second supervisor, Dr. E. McCruden, for her support and advice throughout the project. I appreciate the precious time spent with Dr. R. Bowden. He introduced and taught me network analysis of sequences, and enhanced my ability to understand phylogenetic trees and networks. I thank Prof. T. Schulz for his excellent ideas, discussions and provision of facilities when called upon. I thank Julie Sheldon for running the serological tests and Pam Cook for ideas on K1 PCR. Kathleen Wright did most of the sequencing in these studies, but other people, including Aidan Dolan and Moira Watson, also helped. For this, I am

very grateful. I thank the washroom and media staff for an endless supply of glassware and culture media.

I dedicate this thesis to my children, Peter, Michael and Catherine, who have endured long spells of mummy's absence from home. They have given me all the support during the time they have been in Glasgow including helping out in the house and praying for me to complete the PhD successfully. I owe them a debt that I shall never be able to repay - I love you so much. My sister, Rita, has also been of tremendous help and support. I thank all friends, Parvis - for the wonderful fellowship and meals, her husband Tipu, Wei-Li, Poonam, Steve, Gibson, Gerald, Jessica and husband Chris, Ruth, members of Glasgow SDA church and all others who have contributed in any way to make Glasgow a pleasant place to live in. I thank Charles for being a great lab father. God bless you all.

The author was a recipient of a Commonwealth Academic Staff Scholarship and I thank the Commonwealth Scholarship Commission for their understanding in times of need. I am also indebted to Dr. Joe Conner for his big heart.

Unless otherwise stated, the work presented in this thesis was by the author's own efforts.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

°C	degrees Celsius
µg	microgram
µM	micromolar
aa	amino acid
A	adenine
AIDS	acquired immune deficiency syndrome
bp	base pair
C	cytosine
DNA	deoxyribonucleic acid
dNTPs	deoxyribonucleoside triphosphates
EDTA	ethylenediaminetetra-acetic acid
ELISA	enzyme-linked immunosorbent assay
EtBr	ethidium bromide
g	gram
G	guanine
h	hour
Hb	haemoglobin
IPTG	isopropyl-β-D-thiogalactoside
kbp	kilobase pair
l	litre
M	molar
mg	milligram
min	minute
ml	millilitre
mM	millimolar
nm	nanomolar
nt	nucleotide
ORF	open reading frame
PBMC	peripheral blood mononuclear cells
PCR	polymerase chain reaction
PEG	polyethylene glycol
s	second
TR	terminal repeat
UV	ultraviolet
WBC	white blood cells
w/v	weight/volume
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

CHAPTER 1
INTRODUCTION

The study reported here involved two principal areas: determination of the seroprevalence of human herpesvirus 8 (HHV-8) in Ugandan blood donors and characterization of the genomes of HHV-8 strains in Ugandan Kaposi's sarcoma (KS) patients. This chapter gives an overview of the family *Herpesviridae*, the biology of HHV-8 and KS, as well as the justification and objectives of the study.

1.1. FAMILY *HERPESVIRIDAE*

1.1.1. Overview

The family *Herpesviridae* is a diverse collection of over 120 large DNA viruses infecting all vertebrates, including humans (Minson et al., 2000), and at least one documented invertebrate, the oyster (Comps and Cochennec, 1993). Herpesviruses have a characteristic morphology consisting of both symmetric and nonsymmetric components (Homa and Brown, 1997; Steven and Spear, 1997). The virion (e.g. of herpes simplex virus type 1 [HSV-1]) is spherical, comprising a core, capsid, tegument and envelope (Fig. 1.1). The core consists of the viral genome packed in a liquid crystalline array that fills the entire volume of the preformed icosahedral capsid. The mature capsid is 125 nm in diameter and is composed of 162 capsomeres. The capsid is enclosed within a proteinaceous matrix, the tegument, which in turn is intimately associated with the lipid bilayer, the envelope. The envelope contains a number of different integral glycoproteins (Fig. 1.1). The virions range in diameter between 160 nm and 230 nm, averaging 180 nm (Szilagyi and Berriman, 1994).

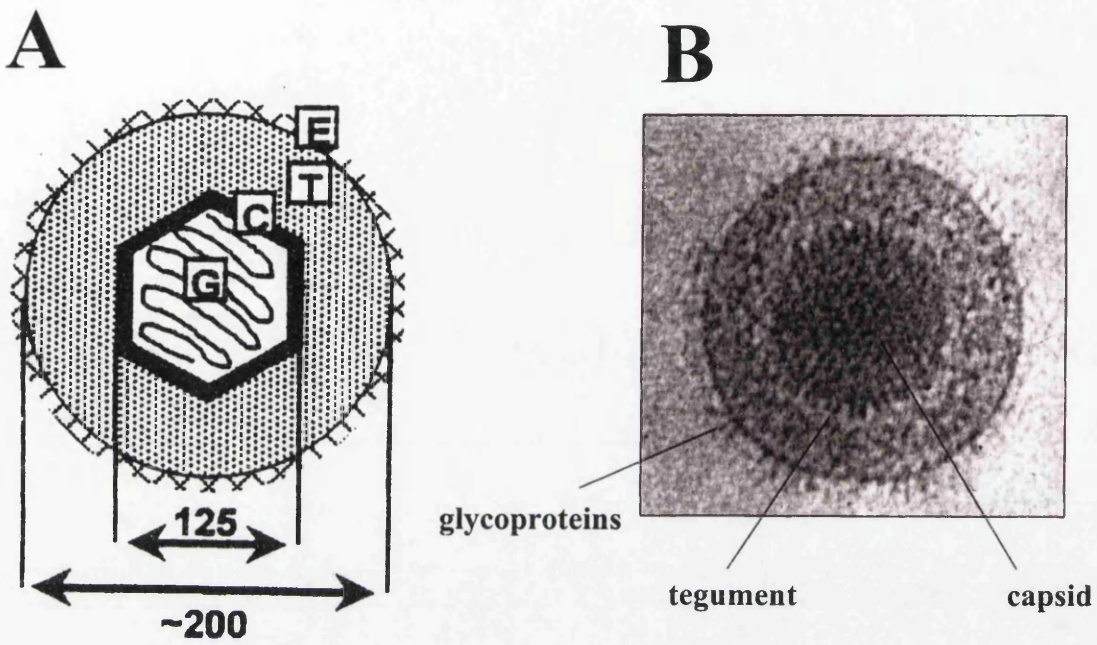


Fig. 1.1. Herpesvirus morphology.

A. Schematic representation of a herpesvirus virion with diameters in nm. E, envelope; T, tegument, C, capsid; G, genome. B. Electron micrograph image of a HSV-1 strain 17+ virion. (Courtesy of F. Rixon).

Herpesvirus replication involves two stages, lytic and latent. Lytic infection follows viral adsorption to the cell, penetration of the nucleocapsid and migration to the nucleus, where gene expression occurs from the viral genome in a sequentially ordered cascade involving three main phases: immediate early, early and late. Transcription, viral DNA replication, capsid assembly, DNA packaging and at least one stage of envelopment take place in the nucleus. New viral progeny are released from the cell by exocytosis.

Following primary infection, all herpesviruses have the ability to persist in an inapparent (or latent) form for the lifetime of the host, with only a subset of viral genes being expressed (Kieff and Liebowitz, 1990; Rock, 1993). Reactivation (due to poorly defined stimuli) occasionally occurs leading to lytic viral production.

Herpesvirus genomes consist of single molecules of linear double-stranded DNA. Those sequenced show a wide range in size: the smallest 125 kbp (SVV; virus names are listed in Table 1.1) and the largest 245 kbp (HCMV, clinical isolates) (McGeoch and Davison, 1999a). The viruses also exhibit an impressively wide range of base compositions ranging from 32-75% in G+C content (Minson et al., 2000). Characteristically, herpesvirus genomes contain terminal or internal repeated sequences, in direct or inverse orientation. The layouts of these repeat elements in relation to unique sequences define at least six genome types (Fig. 1.2). The genomes encode approximately 70 (e.g. HSV-1) to 200 (e.g. HCMV) genes. About 40 of these genes have counterparts in all three subfamilies (*Alpha-*, *Beta-*, and *Gammaherpesvirinae*; see below) and are referred to as core or conserved genes (McGeoch and Davison, 1999a). Most of these genes are involved in vital aspects of the viral life cycle, such as entry into the cell, viral DNA replication, packaging of the genome and virion assembly. Nine genes (of which all but one are involved in DNA metabolism and

Table 1.1. Herpesvirus classification and accession numbers for sequenced genomes

Common name/abbreviation[a]	ICTV[b] designation/abbreviation	GenBank accession no.
ALPHAHERPESVIRINAE		
Simplexvirus (α_1)		
Herpes simplex virus type 1/(HSV-1)	*Human herpesvirus 1*/HHV-1	X14112
Herpes simplex virus type 2/HSV-2	*Human herpesvirus 2*/HHV-2	Z86099
Bovine mamillitis virus/BHV-2	*Bovine herpesvirus 2*/BoHV-2	Not available
Varicellovirus (α_2)		
Varicella-zoster virus/(VZV)	*Human herpesvirus 3*/HHV-3	X04370
Simian varicella virus/SVV	*Cercopithecine herpesvirus 9*/CeHV-9	AF275348
Infectious bovine rhinotracheitis virus/BHV-1	*Bovine herpesvirus 1*/BoHV-1	AJ004801
Pseudorabies virus/PRV	*Suid herpesvirus 1*/SuHV-1	Not available
Feline rhinotracheitis virus/FHV-1	*Felid herpesvirus 1*/FeHV-1	Not available
Equine abortion virus/EHV-1	*Equid herpesvirus 1*/EHV-1	M86664
Equine rhinopneumonitis virus/EHV-4	*Equid herpesvirus 4*/EHV-4	AF030027
"Marek's disease-like virus" (α_3)		
Marek's disease virus 1/(MDV-1)	*Gallid herpesvirus 2*/GaHV-2	AF243438
Marek's disease virus 2/MDV-2	*Gallid herpesvirus 3*/GaHV-3	AB049735
Herpesvirus of turkeys/HVT	*Meleagrid herpesvirus 1*/MeHV-1	AF291866
"Infectious laryngotracheitis-like viruses" (α4)		
Infectious laryngotracheitis virus/(ILTV)	*Gallid herpesvirus 1*/GaHV-1	Not available
BETAHERPESVIRINAE		
Cytomegalovirus (β_1)		
Human cytomegalovirus/(HCMV)	*Human herpesvirus 5*/HHV-5	X17403
Chimpanzee cytomegalovirus/CCMV	*None*/None	Not available
Rhesus cytomegalovirus/RHCM	*Cercopithecine herpesvirus 8*/CeHV-8	Not available
Muromegalovirus (β_1)		
Mouse cytomegalovirus/(MCMV)	*Murid herpesvirus 1*/MuHV-1	U68299
Rat cytomegalovirus/RCMV	*Murid herpesvirus 2*/MuHV-2	AF232689
Roseolovirus (β_2)		
Human herpesvirus 6/(HHV-6)	*Human herpesvirus 6*/HHV-6	X83413, AB021506, AF157706
Human herpesvirus 7/HHV-7	*Human herpesvirus 7*/HHV-7	U43400, AF037218
Unassigned member		
Guinea pig cytomegalovirus/GCMV	*Caviid herpesvirus 2*/CavHV-2	Not available
GAMMAHERPESVIRINAE		
Lymphocryptovirus (γ_1)		
Epstein-Barr virus/(EBV)	*Human herpesvirus 4*/HHV-4	M35547, V01555, M80517
Rhadinovirus (γ_2)		
Herpesvirus saimiri/(HVS)	*Saimiriine herpesvirus 2*/SaHV-2	X64346
Herpesvirus ateles/HVA	*Ateline herpesvirus 2*/AtHV-2	AF083424
Kaposi's sarcoma-associated herpesvirus/KSHV	*Human herpesvirus 8*/HHV-8	U75698, U93872
Rhesus rhadinovirus/RRV	*Cercopithecine herpesvirus 17*/CeHV-17	AF083501, AF210726
Mouse herpesvirus 68/MHV68	*Murid herpesvirus 4*/MuHV-4	U97553, AF105037
Equine herpesvirus 2/EHV-2	*Equid herpesvirus 2*/EHV-2	U20824
Malignant catarrhal fever virus/AHV-1	*Alcelaphine herpesvirus*/AlHV-1	AF005370
Bovine herpesvirus 4/BHV-4	*Bovine herpesvirus 4*/BoHV-4	AF318573
UNDEFINED SUBFAMILY		
"Ictalurid herpes-like virus"		
Channel catfish virus/(CCV)	*Ictalurid herpesvirus 1*/IcHV-1	M75136

Footnotes overleaf.

Footnotes for Table 1.1

[a] Genera names and synonyms (lowercase bold) are given for each subfamily (uppercase bold). Type species are enclosed in parentheses.

[b] ICTV, International Committee on Taxonomy of Viruses (Minson et al., 2000).

Adapted from Davison (2001).

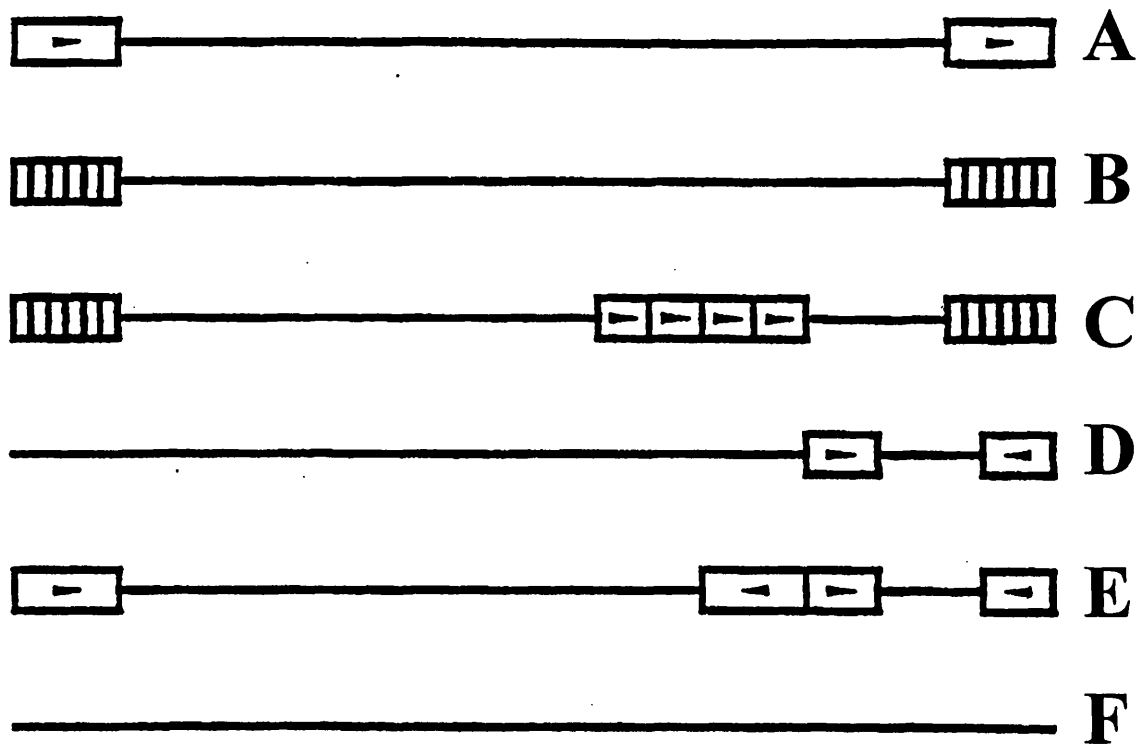


Figure 1.2. Types of herpesvirus genome structure.

Unique sequences and repeat elements are represented by single lines and rectangles, respectively (not drawn to scale). Relative orientation for large repeats are indicated by arrowheads. The number of terminal repeats in types B and C, and the number of internal repeats in type C, are variable. Adapted from McGeoch & Davison (1999a).

replication) are clearly related by similarities in encoded amino acid sequences to cellular genes (McGeoch and Davison, 1999a), and probably originated from the cellular genome. Viral structural genes, on the other hand, do not appear to have any obvious cellular homologues. Most genes involved in control of the life cycle, latency and cell-virus interactions are not conserved in the three subfamilies (Davison, 2001).

The gene complements of herpesviruses have diverged by acquisition of cellular genes and by other recombination processes, including gene duplication and rearrangement, and possibly by genes arising *de novo* (e.g. US11 gene of HSV-1; Rixon and McGeoch, 1984). The genes have also evolved via extensive nucleotide mutations (substitutions, insertions and deletions).

1.1.2. Classification and phylogeny of herpesviruses

The herpesviruses that infect mammals and birds are grouped into three subfamilies, the α-herpesviruses (formally known as the *Alphaherpesvirinae*), the β-herpesviruses (*Betaherpesvirinae*) and the γ-herpesviruses (*Gammaherpesvirinae*), based on biological characteristics (Roizman et al., 1992, 1995). The α-herpesviruses exhibit rapid cytolytic growth in vitro and establish latency in the nervous system. The β-herpesviruses have a long productive cycle, usually associated with the formation of enlarged (cytomegalic) cells, and have a restricted host range in vitro. The γ-herpesviruses establish latent infection in lymphocytes and are often associated with lymphoproliferative disease.

Extensive sequence data on herpesviruses have allowed a more objective subdivision of the family based on genomic attributes of the viruses, which correlates well in general with that established previously (McGeoch and Cook,

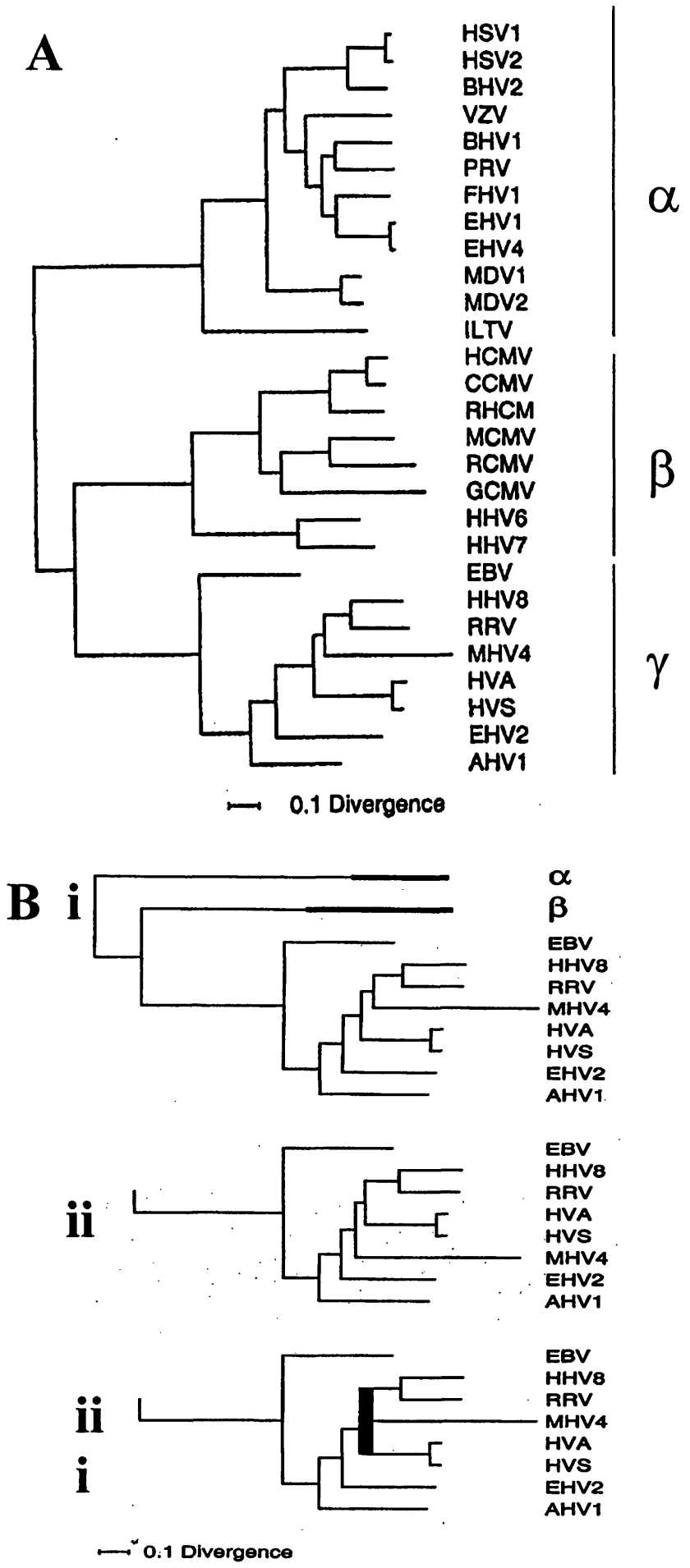
1994; McGeoch et al., 1995, 2000). Table 1.1 gives an outline of the current herpesvirus classification and the accession numbers for the sequenced genomes. Limited sequence data of herpesviruses isolated from turtles (Quackenbush et al., 1998) suggest that these viruses belong to the α-herpesviruses. However, herpesviruses of fish (Davison, 1992, 1998), amphibians (Davison et al., 1999) and one invertebrate (the oyster) (A.J. Davison, unpublished data) are only remotely related to the mammalian/avian group.

An in-depth phylogenetic analysis of mammalian and avian herpesviruses based on a set of eight conserved genes has recently been accomplished by McGeoch et al. (2000) (Fig. 1.3A). This analysis, which included more virus species and used more rigorous tests than previous studies (McGeoch and Cook, 1994; McGeoch et al., 1995), confirmed the presence of an overall high level of congruence between virus hosts' lineages and the herpesvirus phylogenetic tree. This supports the inference made previously that many herpesviruses have evolved with their hosts (cospeciation). Based on this hypothesis and using the divergence times of the host lineages (Benton, 1990; Kumar and Hedges, 1998), it is estimated that the three herpesvirus subfamilies arose approximately 180-220 million years ago, that the genera diverged 80-60 million years ago (probably before the mammalian radiation) and that the species arose within the last 80 million years.

The study by McGeoch et al. (2000) revealed complex features about the phylogeny of the γ-herpesviruses compared with the α- and β-herpesviruses. MHV4 (MHV68) has a uniquely long terminal branch (probably indicating that this lineage has been evolving at a higher rate than other species) and its exact phylogenetic locus could not be resolved (Fig. 1.3B). It appears in various tree locations with different genes. A cospeciational history was also not indicated for this species. The EBV lineage, on the other hand, has an unusually short terminal

Fig. 1.3. Herpesvirus phylogenetic trees.

(A) A maximum likelihood tree based on amino acid sequences of the UL27-plus-UL30 (HSV-1 homologues) gene set of 28 herpesvirus species. (B) A maximum likelihood tree showing γ-herpesviruses; the α- and β-herpesviruses branches are each represented as compressed to a single heavy line in (i). i, the top scoring tree; ii, the second top scoring tree; iii, representation of the uncertain branching order of MHV4 (MHV68) as a heavy bar. See Table 1.1 for full names of the viruses; MHV4 is MHV68. (A) and (B) adapted from McGeoch et al. (2000) and McGeoch (2001), respectively.


A
HSV1
HSV2
BHV2
VZV
BHV1
PRV
FHV1
EHV1
EHV4
MDV1
MDV2
ILTV
HCMV
CCMV
RHCM
MCMV
RCMV
GCMV
HHV6
HHV7
EBV
HHV8
RRV
MHV4
HVA
HVS
EHV2
AHV1
α
β
γ
0.1 Divergence
B i
α
β
EBV
HHV8
RRV
MHV4
HVA
HVS
EHV2
AHV1
ii
EBV
HHV8
RRV
HVA
HVS
MHV4
EHV2
AHV1
ii
i
EBV
HHV8
RRV
MHV4
HVA
HVS
EHV2
AHV1
0.1 Divergence

branch, indicating that it may have a lower evolutionary rate. Next, whereas cospeciation is powerfully evident in the tree for the α- and β-herpesviruses, it is discernible but less obvious in the γ-herpesvirus tree. The branching of Old (HHV-8, RRV) from New (HVA, HVS) World primate viruses (Fig. 1.3) is consistent with cospeciation, but the magnitude of divergence between these two groups implies a rate of change about twice that seen for the α- and β-herpesviruses, another indication of more rapid evolution in the γ_2-herpesviruses. Finally, unlike the ungulate viruses in the α-herpesviruses (EHV1, EHV4), ungulate viruses in the γ-herpesvirus group (AHV1, EHV2) do not form a clade as might have been expected. It was not clear whether this finding is genuine or an artifact of tree inference associated with aberrant substitution rates McGeoch et al., 2000).

1.1.3. Human herpesviruses

Presently, eight herpesviruses, representing all the major mammalian lineages (α_1, α_2, β_1, β_2, γ_1, γ_2; Table 1.1), are known to infect the human species. These viruses cause a wide variety of diseases and possess varied molecular properties (shown in Table 1.2). The infections range from inapparent to severe, disabling or fatal infections in the foetus, the very young or the immunosuppressed. Most of these viruses occur at high prevalence levels in human populations, although HSV-2 is characteristically sexually transmitted and HHV-8 occurs at high prevalence only in some populations (discussed in section 1.8 below). The viruses establish a life long infection in their host, utilizing varied mechanisms of latency; HSV-1, HSV-2 and VZV are neurotropic, HHV-6, HHV-7 and EBV are lymphotropic, while HCMV establishes latency in the monocytic lineage. The situation with HHV-8 will be discussed below. EBV and HHV-8 have involvement with human cancers.

Table 1.2. Human herpesviruses

Virus	Lineage	Genome size (kbp)	G+C Content (%)	Genome type[a]	Encoded proteins[b]	Pathology
HSV-1	α1	152	68	E	74	Recurrent epithelial lesions; rarer, serious neural disease
HSV-2	α1	155	70	E	74	As for HSV-1
VZV	α2	125	46	D	69	Primary infection chickenpox; recurrence as shingles
HCMV	β1	245[c]	57[c]	E	227[c]	Congenital abnormalities; severe infections in immunocompromised
HHV-6	β2	159	43	A	85[d]	Primary infection exanthema subitum
HHV-7	β2	153	36	A	84[d]	No definitive disease association
EBV	γ1	184[e]	60[e]	C	83[e]	Primary infection mononucleosis; associated with Burkitt's lymphoma and other neoplasias
HHV-8[f]	γ2	141	53	B	81	Associated with Kaposi's sarcoma and other neoplasias

[a] See Fig. 1.2.

[b] Estimates of encoded proteins are not precise, as discussed by McGeoch and Davison (1999).

[c] Data are inclusive of regions absent from the sequence for strain AD169 (Cha et al., 1996; Chee et al., 1990; Dargan et al., 1997).

[d] Estimates for numbers of encoded proteins are from Megaw et al. (1998).

[e] Data are inclusive of a region absent from the sequence for strain B95-8 (Parker et al., 1990).

[f] Genome size and G+C content excludes terminal repeats.

Adapted from McGeoch and Davison (1999a).

1.2. DISCOVERY AND CLASSIFICATION OF HHV-8

For a long time, the distribution of KS in the human population suggested involvement of an infectious agent (Beral, 1991, Beral et al., 1990; Oettle, 1962). The eighth, and most recently discovered, human herpesvirus (HHV-8), also referred to as Kaposi's sarcoma-associated herpesvirus (KSHV), is proposed to be this agent (Chang et al., 1994; Ganem, 1996). Employing a PCR-based method (representational difference analysis, RDA), HHV-8 was identified as two DNA fragments (330 and 631 bp in length) in AIDS-KS lesions. These fragments showed high sequence homology to capsid and tegument protein genes, respectively, of two known γ-herpesviruses, HVS and EBV. Subsequent characterization of a larger genomic region containing blocks of structural genes found in all herpesviruses indicated that HHV-8 is a γ_2-herpesvirus (Table 1.1).

HHV-8 is the first human γ_2-herpesvirus to be identified and was originally found to be most closely related to HVS (Moore et al., 1996b). However, it is now known to be more closely related to RRV (Fig. 1.3A), a virus recently isolated from rhesus monkeys (Desrosiers et al., 1997). Several other new HHV-8-related γ_2-herpesviruses (for which only the sequence of a DNA polymerase gene fragment is available) have also been detected in lower and higher Old World primates (Fig. 1.4). The presence of HHV-8-like viruses in primates, especially in the great apes, raises the possibility that these animals could serve as reservoirs of new γ_2-herpesviruses for humans (Lacoste et al., 2000c). It is evident from Fig. 1.4 that two distinct HHV-8-related viruses (the RRV and HHV-8 lineages) are found in the lower primates, while only the HHV-8 lineage has thus far been found in the higher primates (chimpanzees, gorillas and humans). If viruses belonging to the RRV lineage are also present in the higher primates, then an additional HHV-8-like virus may be found in the humans.

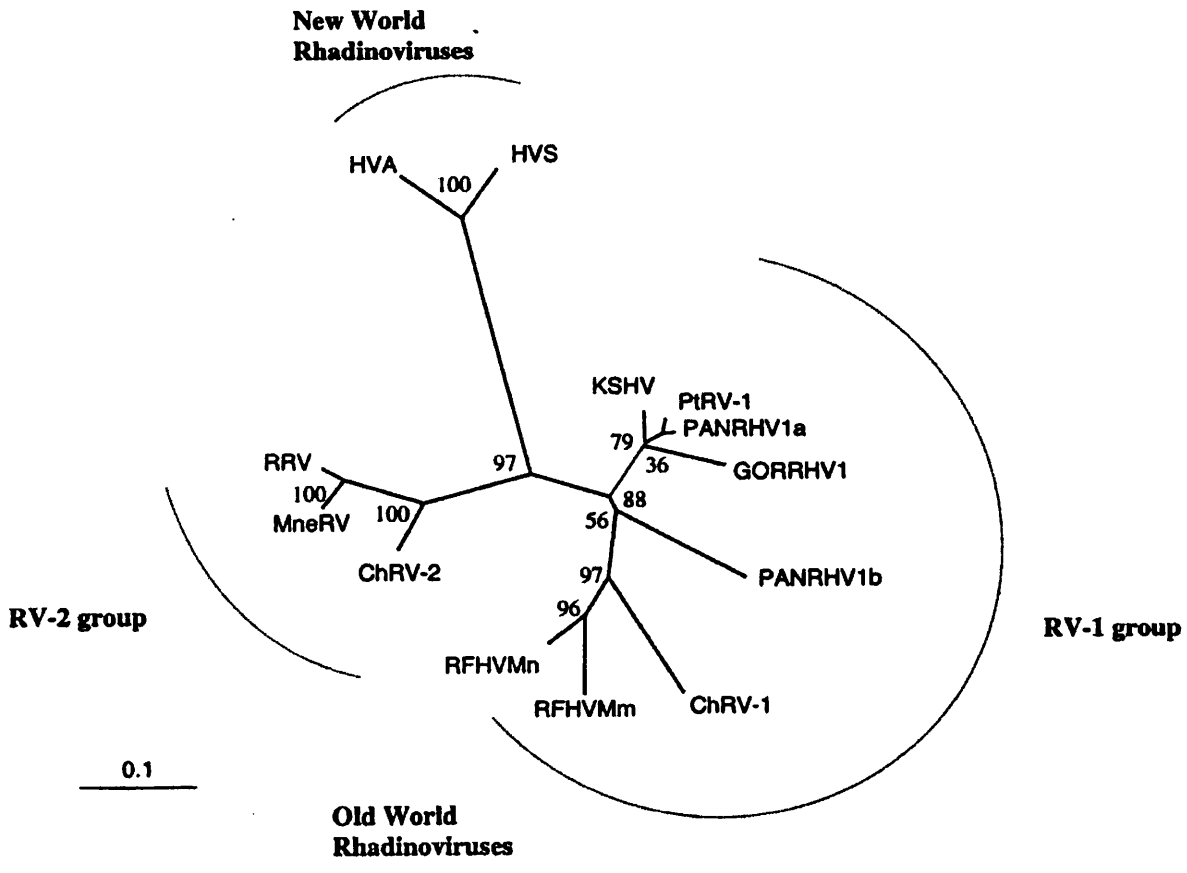


Fig. 1.4. HHV-8-related rhadinoviruses.

A neighbour joining tree based on a 150 amino acid segment of the DNA polymerase gene. The figure illustrates the two lineages of the Old World rhadinoviruses (HHV-8/KSHV/RV-1 and RRV/RV-2). PtRV-1, PANRHV1a and PANRHV1b are from chimpanzees); GORRHV1 is from gorillas; ChRV-1 and ChRV-2, Chlorocebus rhadinovirus 1 and 2, respectively, of African Green monkeys (*Chlorocebus aethiops*); RFHVMm and RFHVMn, retroperitoneal fibromatosis-associated herpesviruses of rhesus (*Macaca mulatta*) and pigtail (*Macaca nemestrina*) macaques; MneRV, was also detected in *M. nemestrina*. RRV infects rhesus macaques. Reproduced from Greensill & Schulz (2000).

1.3. ASSOCIATION OF HHV-8 WITH DISEASE

HHV-8 is associated with three tumour conditions, Kaposi's sarcoma (KS), primary effusion lymphoma (PEL), also known as body cavity-based B-cell lymphomas (BCBL) and multicentric Castleman's disease (MCD) (reviewed by Sarid et al., 1999; Schulz, 1998).

1.3.1. Kaposi's sarcoma

Kaposi's sarcoma (KS) is the leading disorder in AIDS patients, but it also occurs at a lower frequency in HIV-negative individuals (Beral, 1991). It is a vascular disorder involving the skin and mucosal surfaces, and in aggressive cases may involve visceral organs and lymph nodes. The lesion contains spindle cells, activated endothelial cells, fibroblasts, smooth muscle cells, infiltrating inflammatory and lymphoid cells, and an extensive and aberrant neovascularisation (Beckstead et al., 1985; Moses et al., 1999; Roth et al., 1992; Rutgers et al., 1988). Microscopically, KS appears as a disorganised collection of blood-filled vascular slits composed of spindle cells (Fig. 1.5; Sarid et al., 1999). The spindle cells, which possibly originate from the lymphatic or precursor endothelium as opposed to vascular endothelium (Dupin et al., 1999), are thought to play a key role in KS pathogenesis. Whether KS is a true cancer is still debatable. Clonality studies provide evidence for both a monoclonal (neoplasia) and a polyclonal (hyperplasia) origin (Delabesse et al., 1997; Diaz-Cano and Wolfe, 1997; Dupin et al., 1999; Kaaya et al., 1995; Rabkin et al., 1995, 1997). Although KS cells are disseminated, it is not clear whether this represents a multicentric or metastatic process (Rabkin et al., 1997). The lesion generally lacks the nuclear pleomorphism characteristic of many aggressive cancers.

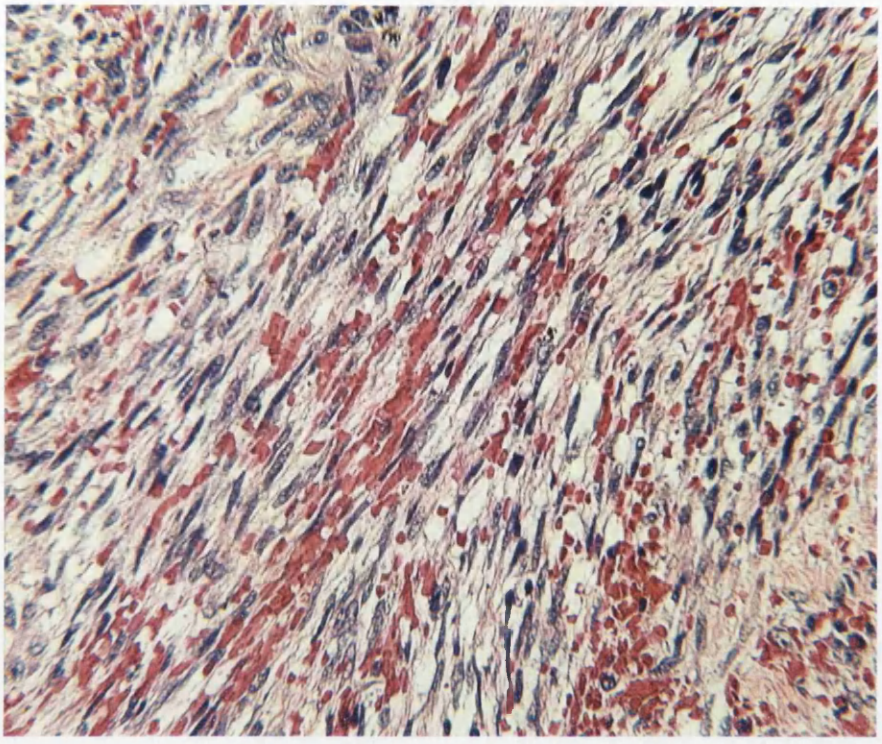


Fig. 1.5. Lung KS lesion (x25 magnification).

Haematoxylin and eosin stain showing the highly vascularized tissue and slit-like cells surrounding the aberrant vessels (slide provided by David Blackbourn).

KS appears to be primarily a growth factor disorder (Ganem, 1996; Levy and Ziegler, 1983). Direct assay of AIDS-KS cells reveals expression of a wide array of cytokines and growth factors, including IL-1, IL-6, basic fibroblast growth factor (bFGF), platelet-derived growth factor (PDGF) and transforming growth factor (TGF)-beta. Some of these factors can stimulate spindle cell growth (Ensoli et al., 1989; Ganem, 1996; Miles et al., 1990). Basic FGF is a particularly powerful angiogenic factor and has been proposed to be an important determinant of the new vascular tissue formation seen in KS (Ensoli et al., 1994). Thus, the following paradigm can be envisaged for the genesis of KS. Spindle cells or their precursors are triggered to proliferate by exogenous factors, such as infectious agents, and HHV-8 could be such a factor. Once activated they produce angiogenic and other factors that induce the vascular and inflammatory components characteristic of the tumour (Ganem, 1996). It is becoming increasingly clear that human immunodeficiency virus type 1 (HIV-1) plays an enhancing role in the pathogenesis of KS, and not a direct etiologic role (Ganem, 1996). HIV-1 appears to increase the risk of KS beyond that expected to arise from immunodeficiency *per se* (Beral et al., 1990). HIV-1 tat protein is known to be a strong angiogenic factor, and can act in synergy with bFGF in the induction of KS (Ensoli et al., 1994).

KS occurs in four forms, based on clinical and epidemiological differences (reviewed by Tappero et al., 1993): classic, African endemic, AIDS/HIV-associated (epidemic) and transplantation-associated. HHV-8 is associated with all four forms (Boshoff et al., 1995b; Buonaguro et al., 1996; Chang et al, 1994, 1996a; Chuck et al., 1996; Dictor et al., 1996; Dupin et al., 1995; Gaidano et al., 1996b; Luppi et al., 1996b; Moore and Chang, 1995; Noel et al., 1996; Schalling et al., 1995; Su et al., 1995).

1.3.1.1. Classic KS

Classic KS, which was first described by Mortiz Kaposi in 1872, occurs predominantly in elderly men of Mediterranean descent. The disease occurs as a benign tumour affecting mainly the skin of the distal extremities (reviewed by Sarid et al., 1999). Classic KS is more common in men than in women probably due to hormonal differences (Fotsis et al., 1994) or differences in other risk factors for HHV-8 infection. HIV-negative gay men may also be affected by classic KS, and they appear to be at a higher risk compared to HIV-negative heterosexual men (Fenoglio et al., 1982; Friedman-Kien et al., 1990; Hjalgrim et al., 1996).

1.3.1.2. African endemic KS

African endemic KS occurs in Central and East African countries, including the Democratic Republic of Congo, Uganda and Zambia, with a particularly high incidence. The disease was already common prior to the AIDS epidemic (Beral, 1991; Taylor et al., 1971). Like classic KS, African endemic KS occurs more commonly in men than in women and, in adults, has a clinical presentation similar to that of classic KS (Fig. 1.6A). African endemic KS patients have no apparent immunosuppression (Kestens et al., 1985), although one study in Tanzania reported some decreased reactivity on antigen skin testing and low CD4+ lymphocyte levels in patients compared to controls (Urassa et al., 1998). Endemic KS also affects children at a median age of 4 years (Ziegler and Katongole-Mbidde, 1996).

1.3.1.3. AIDS/HIV-associated KS

AIDS/HIV-associated (epidemic) KS is the most recently described form of KS and is associated with HIV-1 infection. It was first recognised in the United States in the early 1980s, but is now widely prevalent. The presence of HIV in Africa led to an enormous increase in the incidence of KS, including endemic KS

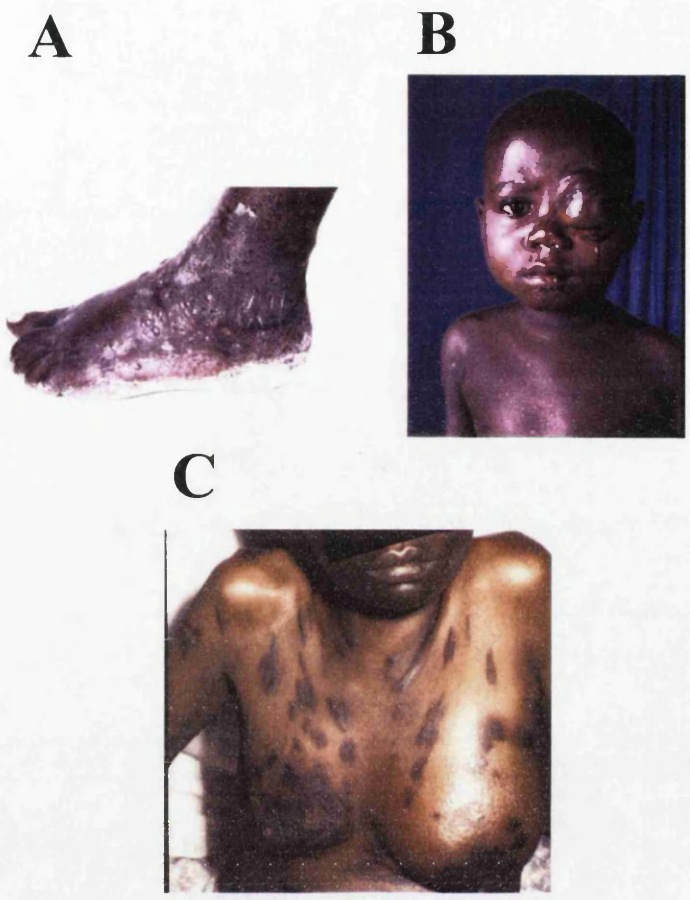


Fig. 1.6. Ugandan KS cases.

A. Endemic KS in an elderly man showing nodular lesions. B. Childhood KS of the oro-facial dominant type (left eye blinded). C. Epidemic (AIDS-associated) KS showing patches. The whole body surface of this patient was affected. Slides A and B were provided by Edward Katongole-Mbidde; slide C was produced with the help of Naomi Byabazaire.

(Athale et al., 1995; Bouquety et al., 1989; Wabinga et al., 1993; Ziegler and Katongole-Mbidde, 1996). For example, childhood KS increased more than 40-fold in Uganda (Ziegler and Katongole-Mbidde, 1996) and approximately 80% of the cases are HIV-associated (E. Katongole-Mbidde, personal communication). KS in children is characterised by a rapidly progressive lymphadenopathy that is generally fatal within 1 year of onset (Ziegler and Katongole-Mbidde, 1996). In addition to the lymphadenopathy, muco-cutaneous lesions are present and the distribution of the tumours shows two major patterns: oro-facial dominant (in the majority of cases; Fig. 1.6B) and inguinal-genital dominant (in the minority of cases). KS has been seen in infants as young as 1 year of age in Uganda (Ziegler and Katongole-Mbidde, 1996), and 7 months of age in Zambia (Athale et al., 1995). Childhood KS is extremely rare outside Africa. Adult AIDS-KS is the most clinically aggressive form of adult KS, most likely as a result of immunodeficiency and the HHV-8 enhancing effect of HIV-1 infection. Internal organs, as well as the whole body surface, are often affected (Fig. 1.6C) (Beral, 1991).

1.3.1.4. Transplantation-associated KS

The occurrence of posttransplantation KS is highly related to iatrogenically induced immunosuppression. It also tends to be clinically aggressive; however, remission has been reported following discontinuation of the immunosuppressive therapy (Besnard et al., 1996; Harwood et al., 1979; Penn, 1983).

1.3.2. Primary effusion/body cavity-based B-cell lymphoma

HHV-8 is also consistently found in all forms of primary effusion/body cavity-based B-cell lymphoma (PEL/BCBL) (Ansari et al., 1996; Carbone et al., 1996; Cesarman et al., 1995a; Gaidano et al., 1996b; Gessain et al., 1997; Komanduri et al., 1996; Nador et al., 1996; Otsuki et al., 1996; Pastore et al., 1995). This is a

rare form of B-cell lymphoma that occurs most frequently in AIDS patients. HHV-8 is also found in a few non-AIDS-related PEL cases (Carbone et al., 1996; Hermine et al., 1996; Nador et al., 1995; Strauchen et al., 1996). PEL is a truly neoplastic disorder characterised by malignant effusions (as opposed to solid growths) of B cells in the pleural or abdominal cavity, immunoglobulin gene rearrangement, lack of most cell surface markers and, unlike Burkitt's lymphoma, a lack of *c-myc* rearrangements (Dupin et al., 1999; reviewed by Jaffe, 1996). The majority of cases of PEL are persistently coinfected with EBV and HHV-8 (Cesarman et al., 1995a,b; Gessain et al., 1997), but occasional cases restricted to HHV-8 infection have also been reported (Carbone et al., 1996; Cesarman et al., 1996a; Renne et al., 1996b; Strauchen et al., 1996). The HHV-8 genome occurs as multiple (approximately 50-100) copies of circular DNA (episomes) in all PEL lymphoma cells (Cesarman et al., 1995a,b; Dupin et al., 1999). Several HHV-8 persistently (latently) infected B-cell lymphoma cell lines have been established from PEL tumours and these have been invaluable in the study of HHV-8 (Arvanitakis et al., 1996; Cesarman et al., 1995b; Gaidano et al., 1996a; Gao et al., 1996b; Renne et al., 1996b; Said et al., 1996). Like the parent tumours, most of these cell lines are coinfected with HHV-8 and EBV (e.g. BC-1), although a few infected with HHV-8 alone (e.g. BCBL-1 and BCP-1) have also been established.

1.3.3. Multicentric Castleman's disease

HHV-8 is frequently detected in lymph nodes of AIDS patients with multicentric Castleman's disease (MCD), a complex lymphoproliferative disorder characterised by fever, adenopathy, splenomegaly and elevated levels of IL-6 in circulation. MCD occurs in two histological forms, the hyaline-vascular and the plasma cell variant (Peterson and Frizzera, 1993). Patients with the plasma cell variant are at increased risk for KS (Corbellino et al., 1996b). HHV-8 is found

much less frequently in HIV-negative MCD patients (Barozzi et al., 1996; Gessain et al., 1996; Soulier et al., 1995), indicating that HHV-8 is less strongly associated with MCD than with KS and PEL.

1.3.4. Other conditions

HHV-8 has been associated with a number of other conditions, including angioimmunoblastic lymphadenopathy, squamous cell skin carcinoma of transplant recipients, angiosarcoma, cutaneous T-cell lymphoma, bone marrow stromal cells of patients with multiple myeloma and benign monoclonal gammopathy (Gyulai et al., 1996a,b; Luppi et al., 1996a; McDonagh et al., 1996; Rady et al., 1995; Rettig et al., 1997; Sander et al., 1996). However, these findings have not been confirmed by others (Boshoff et al., 1995b, 1996; Chang et al., 1994; Dictor et al., 1996; Jin et al., 1996; MacKenzie et al., 1997; Marcelin et al., 1997; Masood et al., 1997; Olsen et al., 1998c, Parravicini et al., 1997a; Pawson et al., 1996; Tomita et al., 1996; Uthman et al., 1996; Whitby et al., 1997). Recently, HHV-8 infection has also been associated with a non-neoplastic condition, bone marrow failure following organ transplantation (Luppi et al., 2000).

1.4. GENERAL CHARACTERISTICS OF HHV-8

As with other tumour viruses (e.g. papillomaviruses in cervical cancer and EBV in nasopharyngeal carcinoma), the study of HHV-8 has met with certain obstacles. First, the virus cannot be propagated in vitro from KS lesions. However, it can be cultured to high copy numbers in several latently-infected PEL cell lines, and these are the primary tools for HHV-8 isolation, characterization and seroepidemiologic studies. Second, it has been difficult to transmit HHV-8 in vitro. Finally, no animal model is readily available to study

HHV-8 infection. Thus, most work in HHV-8 has employed molecular biological techniques rather than the methods of traditional virology.

A number of cell types (including cells of B-cell, endothelial, epithelial and fibroblastic origin) have been tested for culture of HHV-8. Most support replication either poorly or not at all, and infection is generally only detectable by PCR-based assays (Blackbourn et al., 1997; Foreman et al., 1997; Ganem, 1996; Renne et al., 1998). The adenovirus-transformed human embryonal kidney epithelial 293 cell line was used to culture HHV-8 transiently from KS lesions (Foreman et al., 1997), thus confirming the presence of HHV-8 virions in KS tumour samples. This cell line was also used to isolate HHV-8 from saliva (Vieira et al., 1997).

Latently infected PEL cell lines contain a small proportion of cells undergoing lytic virus production (Orenstein et al., 1997). Lytic replication can be induced in other cells by treatment with sodium butyrate or phorbol esters, such as 12-*O*-tetradecanoylphorbol-13-acetate (TPA) (Lennette et al., 1996; Miller et al., 1996, 1997). Unfortunately, in most cases the supernatants produced by induced cell lines are not highly infectious (Sarid et al., 1999; Schulz, 1998). Recently, a new PEL cell line (JSC-1) that yields a highly infectious HHV-8-containing supernatant has been reported (Cannon et al., 2000). Analysis of virions produced from induced PEL cell lines showed that HHV-8 has characteristics of other herpesviruses (Minson et al., 2000). The HHV-8 capsid is 110 nm in diameter (Sarid et al., 1999) and its three dimensional structure has recently been determined using electron cryomicroscopy and computer reconstruction (Wu et al., 2000). Virions containing DNA are 120-150 nm in diameter (Sarid et al., 1999). Similar particles have also been observed in KS lesions in a few productively infected spindle cells (Ioachim, 1995; Orenstein et al., 1997; Walter et al., 1984). Like other herpesviruses, HHV-8 replicates in the cell nucleus and

undergoes latency, a stage at which its genome is present in the form of circular episomes (Sarid et al., 1999).

The effect of herpesvirus DNA polymerase inhibitors on lytic HHV-8 replication in TPA-induced PEL cell lines provides indirect experimental evidence for the role that HHV-8 may play in KS and PEL pathogenesis. Phosphonoacetic acid (foscarnet), ganciclovir and cidofivir, but not acyclovir, inhibit HHV-8 replication at pharmacological concentrations (Kedes and Ganem, 1997). A number of studies also suggest that foscarnet reduces the rate of KS in AIDS patients (Glesby et al., 1996; Jones et al., 1995a; Mocroft et al., 1996), and, in one small study, it induced the regression of existing lesions (Morfeld and Torssander, 1994). However, neither foscarnet nor ganciclovir eliminates PCR-detectable HHV-8 from the PBMC of infected individuals (Humphrey et al., 1996), indicating that DNA polymerase inhibitors do not eliminate latent infections.

1.5. HHV-8 GENOME AND TUMORIGENESIS

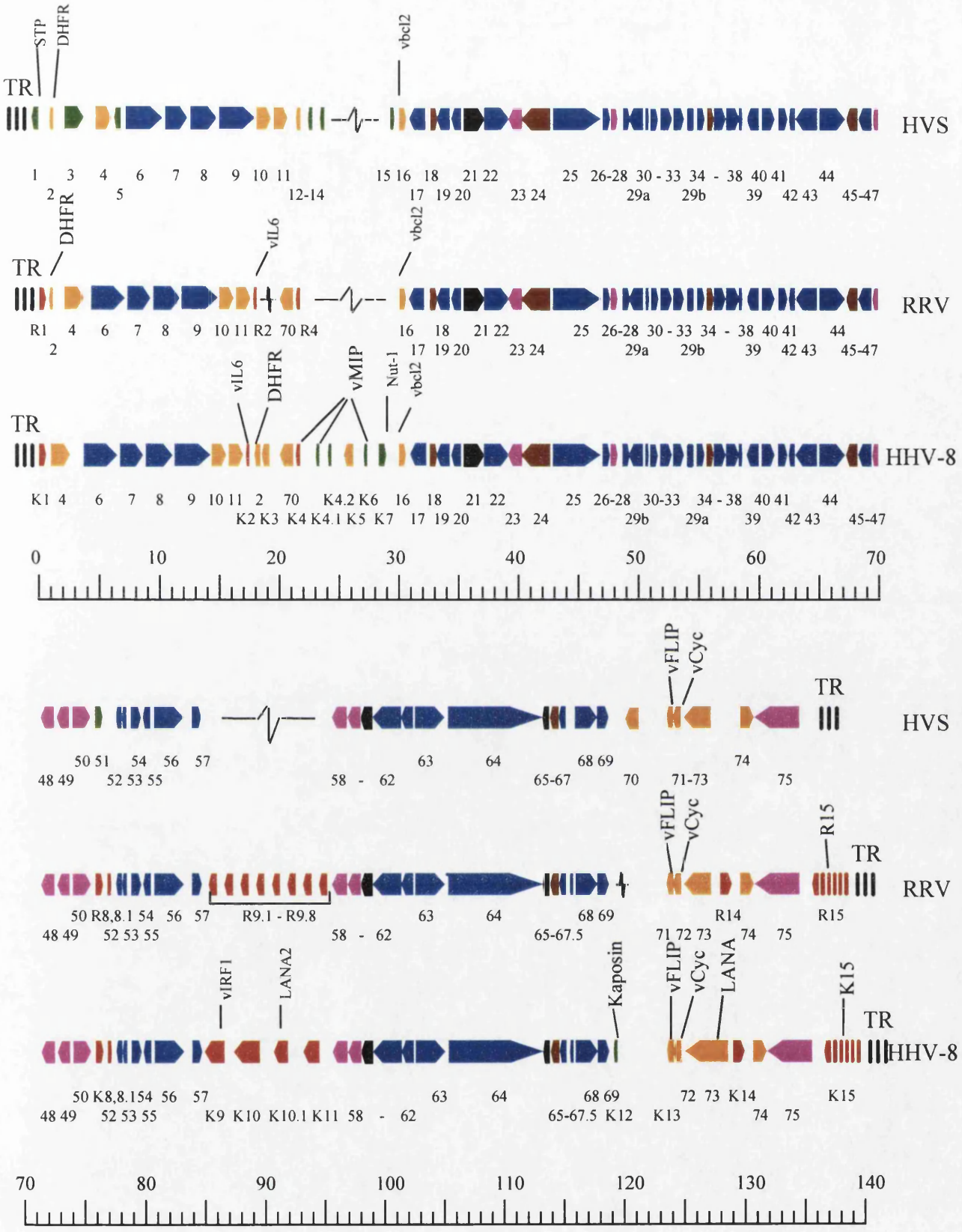
The first published almost complete sequence of the HHV-8 genome (lacking an unclonable G+A-rich repeat segment between the right end of the unique region and the terminal repeat [TR] region) was from the PEL-derived cell line, BC-1 (Russo et al., 1996). Subsequently, another almost complete sequence from an AIDS-KS lesion (lacking part of the coding sequences at the right end of the unique region) was obtained (Neipel et al., 1997a). The size of circular HHV-8 genome in BC-1 cells is 270 kbp on pulse field gel electrophoresis (PFGE) analysis (Moore et al., 1996). In contrast, that of encapsidated linear viral DNA from another cell line (BCBL-1) is 160-170 kbp on PFGE analysis (Renne et al., 1996a; Zhong et al., 1996), a size similar to that of HVS and just smaller than that of EBV. The BC-1 HHV-8 genome is bigger largely due to the presence of a

duplicated part of the unique region of the genome within the TR (Russo et al., 1996). This rearrangement most likely occurred in the parental tumour since it is also found in another cell line (HBL-6) independently derived from the same tumour (Gaidano et al., 1996a).

The organisation of the HHV-8 genome (Fig. 1.7) is very similar to that of other γ_2-herpesviruses, in particular RRV, HVS and MHV68 (Albrecht et al., 1992; Alexander et al., 2000; Neipel et al., 1997a; Nicholas et al., 1997a; Russo et al., 1996; Searles et al., 1999; Virgin et al., 1997; Table 1.1). It consists of a single unique region, about 140.5 kbp in length, flanked by multiple G+C-rich 801 bp tandem repeats (TR), which may be involved in DNA packaging. The total length of the TR region (in circularized DNA) is estimated to vary between 25 and 35 kbp (i.e. 35 to 45 tandem repeat units; Lagunoff and Ganem, 1997). The unique region has a G+C content of 53.5%, and contains all the known protein-coding regions of the HHV-8 genome. At least 81 open reading frames (ORFs) encoding proteins, including 66 homologous to HVS ORFs, were predicted by Russo et al. (1996). It is currently estimated that HHV-8 possesses at least 86 genes (McGeoch and Davison, 1999b). HVS gene nomenclature forms the basis for HHV-8 gene nomenclature (Albrecht et al., 1992; Russo et al., 1996). The ORFs homologous to HVS ORFs (ORF2-75) are assigned the corresponding HVS ORF number, even if their relative position in the HHV-8 genome is different from that in HVS (as is the case with dihydrofolate reductase [ORF2] and thymidylate synthase [ORF70] genes). ORFs categorised as unique to HHV-8 (before the RRV sequence became available) are numbered in consecutive order with a K prefix, i.e., K1 to K15, with newly identified ORFs being assigned a decimal notation (e.g. K4.1). All these genes are now known to have homologues in RRV, except K3, K5, K7 (nut-1) and K12 (kaposin). The K3 and K5 genes (which possibly play a role in immune evasion by downregulation of MHC class 1; Coscoy and Ganem, 2000; Ishido et al., 2000) are homologous to

Fig. 1.7. HHV-8 genome in comparison with HVS and RRV.

Open reading frames are represented by boxes with the arrow ends indicating the direction of transcription. Colour shading: blue, core herpesvirus genes; orange, genes found in γ_2-herpesviruses only; pink, genes found in γ-herpesviruses only; black, genes found in α- and γ-herpesviruses; brown, genes found in β- and γ-herpesviruses; red, genes found in HHV-8 and RRV only; green, genes unique to a particular virus. Adapted from Greensill & Schulz (2000) and amended on the basis of data in Alexander et al. (2000).


TR
STP
DHFR
vbcl2
HVS
1 3 4 6 7 8 9 10 11 15 16 18 21 22 25 26-28 30 - 33 34 - 38 40 41 44
2 5 12-14 17 19 20 23 24 29a 29b 39 42 43 45-47
TR
DHFR
vIL6
vbcl2
RRV
R1 4 6 7 8 9 10 11 R2 70 R4 16 18 21 22 25 26-28 30 - 33 34 - 38 40 41 44
2 17 19 20 23 24 29a 29b 39 42 43 45-47
TR
vIL6
DHFR
vMIP
Nut-1
vbcl2
HHV-8
K1 4 6 7 8 9 10 11 2 70 K4.2 K6 16 18 21 22 25 26-28 30-33 34 - 38 40 41 44
K2 K3 K4 K4.1 K5 K7 17 19 20 23 24 29b 29a 39 42 43 45-47
0 10 20 30 40 50 60 70
vFLIP
vCyc
TR
HVS
50 51 54 56 57 63 64 68 69 74
48 49 52 53 55 58 - 62 65-67 70 71-73 75
vFLIP
vCyc
R15
TR
RRV
50 R8,8.1 54 56 57 R9.1 - R9.8 63 64 68 69 R14 R15
48 49 52 53 55 58 - 62 65-67.5 71 72 73 74 75
vIRF1
LANA2
Kaposin
vFLIP
vCyc
LANA
K15
TR
HHV-8
50 K8,8.1 54 56 57 63 64 68 69 72 73 K14 K15
48 49 52 53 55 K9 K10 K10.1 K11 58 - 62 65-67.5 K12 K13 74 75
70 80 90 100 110 120 130 140

each other as well as to HVS ORF12 (McGeoch and Davison, 1999b). ORF12 is not present in RRV.

The HHV-8 genome contains blocks of genes that are conserved between herpesvirus subfamilies (Fig. 1.7). The core genes are currently estimated to be 42 in number (McGeoch and Davison, 1999b). Between the conserved gene blocks are clusters of nonconserved genes. A number of these genes have homologues in the cellular genome (Neipel et al., 1997a,b,c; Nicholas et al., 1997a,b, 1998; Russo et al., 1996), and possibly play a role in averting cellular antiviral responses. Examples of these genes include a complement binding protein (ORF4), three CC chemokines (vMIPs-K4, K4.1 and K6), four viral interferon regulatory factors (vIRFs-K9, K10, K10.5 and K11), and three potential cellular growth regulators (vIL6 [K2], vcyclin D [ORF72] and viral G-protein-coupled receptor [vGCR or IL8R-like gene; ORF74]). Viral IL6, the vIRFs, and two other cellular homologues, vbcl2 (ORF16) and vFLIP (K13/ORF71) may be involved in inhibiting apoptosis (Chang et al., 1996a; Djerbi et al., 1999; Moore et al., 1996a; Rivas et al., 2001; Thome et al., 1997). The vcyclin D gene is expressed in KS lesions (Chang et al., 1996a; Cesarman et al., 1996b) and can substitute for human cyclin D in phosphorylating (i.e. inactivating) the retinoblastoma protein (pRB) (Chang et al., 1996a), and thus may play a significant role in KS pathogenesis. Although the vIL6 gene is functional (Chang et al., 1996a; Moore et al., 1996a), it is expressed in only a small number of cells in the KS lesion, and thus it is questionable whether it plays a significant role in the prominent angiogenesis seen in KS (Moore et al., 1996a). The vbcl-2 is also functional (Sarid et al., 1997). There is a striking parallel between the cellular homologues encoded by HHV-8 and the cellular genes known to be induced by EBV infection (Russo et al., 1996; Sarid et al., 1999). Cellular cyclin D, CD21/CR2, bcl-2, an IL8 receptor-like protein (EB11), IL-6, MIP-1a, MIP-1b, RANTES and adhesion molecules are upregulated by

EBV infection (Birkenbach et al., 1993; Finke et al., 1992, 1994; Jones et al., 1995b; Palmero et al., 1993). Thus, by introducing exogenous genes, HHV-8 appears to modify the same signalling and regulation pathways as EBV.

Although the genome of HHV-8 (BC-1 strain) does not contain sequence homologues of the major transforming proteins of HVS (STP and Tip; Jung and Desrosiers, 1991; Jung et al., 1991) or of EBV (EBNA1, EBNA2, EBNA3, LMP1, LMP2 or gp350/220; Kieff and Liebowitz, 1990; Sarid et al., 1999), dysregulation of the cell cycle by the identified HHV-8-encoded proto-oncogenes and cytokines may contribute to tumorigenesis. HHV-8 from induced cell lines has been reported to be able to change the morphology and prolong the survival of some cells, including primary human keratinocytes and endothelial cells (Cerimele et al., 2001; Flore et al., 1998; Moses et al., 1999). A number of HHV-8 genes have been shown to be oncogenic in vitro, including K1 (Lee et al., 1998b; see below), vGCR (Arvanitakis et al., 1997; Bais et al., 1998; Cesarman et al., 1996b; Guo et al., 1997), K12 (kaposin; Muralidhar et al., 1998), and K9 (Gao et al., 1997; Li et al., 1998; Zimring et al., 1998). However, the role of these genes in natural infection is not clear since expression of their proteins has not been demonstrated in HHV-8 latently infected KS or PEL tumours (Katano et al., 2000b; Kirshner et al., 1999; Parravicini et al., 2000; Sarid et al., 1997). Presently, only one protein, the latency-associated nuclear antigen (LANA/LANA1 also referred to as latent nuclear antigen [LNA/LNA1]), encoded by ORF73 (Kellam et al., 1997; Rainbow et al., 1997), has been shown by immunohistochemistry to be expressed in all HHV-8-infected cells (Dupin et al., 1999; Katano et al., 2000b; Parravicini et al., 2000). LANA appears to have a multifunctional role in HHV-8 infection. It is involved in maintenance of latency (by tethering the viral genome to host chromatin during mitosis), inhibition of apoptosis (by targeting p53 and pRB) and regulation of viral and cellular transcription (Ballestas et al., 1999; Friborg et al., 1999; Krithivas et al., 2000; Renne et al., 2001; Schwam et al., 2000). A recent study has also implicated LANA in tumorigenesis. In cooperation with the cellular oncogene *Hras* (Harvey rat sarcoma

viral oncogene), LANA transforms primary rat embryo fibroblasts and renders them tumorigenic in nude mice (Radkov et al., 2000).

1.6. HHV-8 GENE EXPRESSION AND TROPISM

1.6.1. Gene expression

Initial screening for HHV-8 transcripts was performed by Zhong et al. (1996) using hybridisation of viral genomic segments with radiolabeled cDNA probes derived from KS tissue and the BCBL-1 cell line. Two highly abundant transcripts, T0.7 (spanning K12) and T1.1 (also referred to as polyadenylated nuclear (PAN) RNA or nut-1) were identified in this study. T1.1 contains features of both mRNAs and small nuclear RNAs (snRNAs). Although it is polyadenylated and is transcribed by RNA polymerase II, it lacks a trimethylguanosine cap and appears to be a non-coding transcript (Sun et al., 1996; Zhong et al., 1996). It can associate with small ribonuclear proteins and contains regions that are homologous or complementary to U1 and U12 snRNAs (Hall and Padgett, 1996; Tarn and Steitz, 1996), suggesting that it might be involved in RNA processing (Sun et al., 1996; Zhong and Ganem, 1997). T1.1 is expressed in a few tumour cells undergoing lytic replication in the KS tumour and its expression is enhanced by chemical induction of PEL cell lines (Sarid et al., 1998; Zhong et al., 1996). Thus, it differs from non-coding nuclear RNAs encoded by EBV (EBERs), HVS (HSURs), and HSV (LATs), which are expressed as latent transcripts.

T0.7 spans the K12 gene (absent from RRV, see above) which encodes a small hydrophobic protein (kaposin) of 60 amino acid residues (Neipel et al., 1997a; Russo et al., 1996). This protein has been shown to have oncogenic potential in vitro (Zhong et al., 1996). T0.7 is expressed abundantly in KS spindle cells and

in persistently infected PEL cells (Renne et al., 1996b; Staskus et al., 1997; Stürzl et al., 1997a; Zhong et al., 1996). Its expression is also induced by chemicals in PEL cell lines (Sarid et al., 1998).

Sarid et al. (1998) used Northern blot analysis and DNA probes spanning the entire genome to map HHV-8 transcripts in induced and uninduced BC-1 cells. In this study, HHV-8 gene transcription was categorised into three broad classes. Class I genes (the noninduced or "latency" genes) are transcribed constitutively and are not enhanced by chemical induction. This class comprises of three genes, vcyclin (ORF72), vFLIP (ORF71) and LANA (ORF73). Class II genes (the enhanced genes) are transcribed to higher levels in induced cells. These genes are primarily clustered in nonconserved regions of the genome and include T1.1 and T0.7 as well as most of the virus-encoded cytokines and signal transduction genes (e.g. vMIPs, vIL6, vIRF-1 and vIRF-2). It is not clear whether the recently identified vIRF-3 (also known as LANA2) encoded by K10.5/K10.6 (Lubyova and Pitha, 2000; Rivas et al., 2001) is a Class I or Class II gene. One group reported an increase in its expression on TPA treatment (Lubyova and Pitha, 2000), while the other observed no such increase (Rivas et al., 2001). Class III genes (the inducible or "lytic" genes) are transcribed only in induced cells and cluster primarily in the conserved regions of the genome. A small population of cells in KS tissue expresses inducible genes and these are believed to correspond to lytically infected cells (Staskus et al., 1997; Stürzl et al., 1997a,b; Sun et al., 1999). Thus, the lytic cycle of HHV-8 may also directly contribute to tumour pathogenesis.

As in other herpesviruses, HHV-8 gene transcription can be classified into four distinct kinetic stages: latent, immediate early, early and late (Sun et al., 1999). Latent gene expression (e.g. of LANA and vFLIP) is constitutive and unaffected by chemical induction. Immediate early genes (e.g. the transactivator gene, Rta,

encoded by ORF50) are induced within 4 h of chemical treatment and are not inhibited by cycloheximide. Two sets of HHV-8 early genes were described by Sun et al. (1999), those expressed within 8 to 13 h after chemical induction (e.g. K3, K5 and vIL6), and those that are slightly delayed (e.g. vGCR, K12 and vbcl2). Late genes (e.g. the structural gene, ORF25) do not appear until after 30 h after induction and their expression is abrogated by phosphonoacetic acid, an inhibitor of HHV-8 DNA replication.

HHV-8 Rta (a homologue of EBV rta, also called BRLF1 or R) is an immediate early gene that is conserved among all γ-herpesviruses (Liu et al., 2000; Manet et al., 1989; van Santen, 1993; Whitehouse et al., 1997; Wu et al., 2000). It plays a central role in the switch of the viral life cycle from latency to lytic replication. Rta has been shown to be necessary and sufficient to activate lytic gene expression in latently infected cells (Lukac et al., 1998, 1999; Sun et al., 1998).

HHV-8 gene expression in PEL induced and uninduced cells has also been studied using DNA arrays (Jenner et al., 2001). This method allowed the simultaneous measurement of the expression level of almost every known HHV-8 ORF. Cluster analysis was used to group together genes that share similar patterns of expression. Genes that may be involved in a common process and genes in the same kinetic stage of herpesvirus replication group together. Overall the expression patterns determined by HHV-8 DNA array analysis were consistent with those described by Sarid et al. (1998) and Sun et al. (1999). In addition, the study allowed categorization of genes whose expression had hitherto not been analysed (e.g. ORF11 and ORF58), as well as the prediction of possible roles for genes that have yet to be characterized (e.g. K10.7, another homologue of IRFs).

There is evidence to suggest that gene expression may be tissue-specific for the various HHV-8-related disorders. For example, vIL6 is expressed in PEL cells and HHV-8-infected B-cells in lymphatic tissue, but not in KS spindle cells (Moore et al., 1996a; Staskus et al., 1999). Also, LANA2 is expressed in B cells but not in KS spindle cells (Rivas et al., 2001). Moreover, Friborg et al. (1998) showed that HHV-8 isolates in PEL cell lines may be biologically distinct from those in primary KS lesions. The virus derived from KS lesions could be propagated in 293 cells, while propagation of virus derived from PEL cell lines was not observed in these cultures.

1.6.2. Tropism

HHV-8 has been detected in a wide range of cells in infected individuals, including endothelial and spindle cells of KS lesions, haematopoeitic cells (B cells, CD8+ cells and macrophages), and epithelial cells (keratinocytes, oropharyngeal, duodenal and rectal mucosa, and prostatic glandular cells) (Ambroziak et al., 1995; Boshoff et al., 1995a; Corbellino, 1996c; Harrington et al., 1996; Li et al., 1996; Mesri et al., 1996; Pauk et al., 2000; Sirianni et al., 1997; Staskus et al., 1997; Stürzl et al., 1997 a,b; Thomas et al., 1996). LANA immunohistochemistry (Dupin et al., 1999; Rainbow et al., 1997) and *in situ* hybridisation (Staskus et al., 1997; Stürzl et al., 1997a,b) confirm HHV-8 gene expression in the majority (>90%) of nodular KS spindle cells and in <10% of cells forming the walls of the aberrant vessels in early KS lesions (Dupin et al., 1999). Lytic replication of HHV-8 occurs in monocytes/macrophages (Stürzl et al., 1997a) and there is presence of linear viral genomes in PBMCs, suggesting productive infection (Decker et al., 1996).

1.7. VARIATION AND EVOLUTION OF HHV-8 STRAINS

The genome is well conserved within most coding regions between differing HHV-8 strains (Neipel et al., 1997a; Poole et al., 1999; Russo et al., 1996). However, genes (K1 and K15) at both ends of the unique region display striking variability. The K1 gene, located at the left end of the genome, shows high diversity (up to 17.5% nucleotide variation) among strains (Kasolo et al., 1998, Cook et al, 1999; Lacoste et al., 2000a; Meng et al., 1999; Zong et al., 1999). K1 occupies a position equivalent to STP of HVS, LMP1 of EBV and R1 of RRV (Damania et al., 1999; Hayward, 1999). However, no sequence homology is apparent between these genes. Expression of K1 is significantly increased during the early phase of lytic replication, indicating that this gene most likely plays a key role at this stage of the virus life cycle (Lagunoff and Ganem, 1997). K1 encodes a type 1 membrane protein with structural resemblance to R1 but not to STP or LMP1. Transfection experiments showed that this protein is highly glycosylated and is principally associated with the plasma and cytoplasmic membranes (Lee et al., 1998b). K1 has been shown to transform rodent fibroblasts and to possess in vivo tumorigenic potential when inserted into an STP-negative (non-transforming) HVS strain (Lee et al., 1998b). However, its tumorigenic potential in natural infection is still unclear. K1 contains a highly conserved region in its cytoplasmic domain with a sequence similar to that of immunoreceptor tyrosine-based activation motifs (ITAMs) (Lee et al., 1998a). Using a chimera containing the N-terminus of CD8 fused to the C-terminus of K1, Lee et al. (1998a) showed that the putative K1 ITAM is functional and that it can transduce signals to induce cellular activation. The full length K1 protein was also shown to be a constitutive signal transducer in B cells (Lagunoff et al., 1999). Recently, Lee et al. (2000) reported that K1 downregulates BCR surface expression in stably transfected B cells.

The amino acid sequence of the K1 protein shows up to 44% variation, with variability concentrated in two extracellular domains, VR1 and VR2 (Cook et al.,

1999; Lacoste et al., 2000a; Meng et al., 1999; Zong et al., 1999). Variability is driven by positive selection (Cook et al., 1999), an evolutionary phenomenon in which non-synonymous nucleotide changes outnumber synonymous changes, with striking alterations in amino acid sequence. The biological forces behind this phenomenon are not known.

HHV-8 subtypes were originally defined on the basis of limited variability within highly conserved genes, such as ORF25, ORF26 and ORF75 (Kasolo et al., 1998; Luppi et al., 1997; Zong et al., 1997). This method did not always allow clear genotypic or epidemiological discrimination between strains. The extensive polymorphism in K1, on the other hand, has proven to be more phylogenetically robust, allowing the performance of molecular epidemiological studies. Five main HHV-8 K1 subtypes, named A (I), B (IV), C (II), D (III) and E (Table 1.3), have been defined (Biggar et al., 2000; Cook et al., 1999; Meng et al., 1999; Nicholas et al., 1998; Zong et al., 1999). Subtype B appears to predominate in Africa together with a variant (A5) of the A subtype so far seen only in African samples (Cook et al., 1999; Kasolo et al., 1998; Lacoste et al., 2000a; Meng et al., 1999). Subtypes A and C occur in Europe and the USA, subtype C in northern Asia, and subtype D in southern Asia, Australia and New Zealand. Subtype E has thus far been found in Brazilian Amerindians. There is variation within the main K1 subtypes (up to 5%) and this has been used to define subgroups (Table 1.3; Cook et al., 1999; Meng et al., 1999; Poole et al., 1999).

A possible link between specific HHV-8 subtypes and diseases (such as KS, MCD, PEL) has been proposed by some authors implying that some subtypes may have different biological properties,. Boralevi et al. (1998) suggested that subtype A may be responsible for more aggressive KS tumours in France. In addition, studies by Luppi et al. (1997) suggested that subtype A strains predominate in Italian classic KS, whereas subtype C strains may be more

Table 1.3. Nomenclature of HHV-8 K1 subtypes[a]

First (highest) level	Second level	Third level	Reference
Subtypes A-D	Subgroups A1-A5, C1-C5, D1 and D2	Clades C3' and A1'	Zong et al. (1999)
Subtypes A-C	Subgroups A1, A3, A5, large subgroup A', C' and C''	-	Cook et al. (1999)
Genotypes I-IV	Subtypes IA-IF	-	Meng et al. (1999)
Subtype E	-	-	Biggar et al. (2000)

[a] Modified slightly from Meng et al. (2001).

prevalent in Italian non-malignant lymphoproliferations. These data, however, were based on a small fragment of the ORF26 gene, recently shown to be ambiguous for distinguishing between strains (Poole et al., 1999). In contrast, most studies on K1 variability have not found a correlation between HHV-8 K1 subtypes and the various diseases or even with the different clinical and epidemiological forms of KS (Cook et al., 1999; Kasolo et al., 1998; Lacoste et al., 2000a; Meng et al., 1999; Zong et al., 1999). HHV-8 subtypes, however, show an apparent geographic and ethnic restriction (Cook et al., 1999; Meng et al., 1999; Zong et al., 1999). This led Hayward (1999) to propose that HHV-8 branched into its various K1 subtypes within the last 10^5 years, in correlation with the migratory patterns of humans out of Africa (Cavalli-Sforza and Cavalli-Sforza, 1995). It has been proposed that subtype B is older than subtypes A and C, which could have diverged more recently (in the order of 10^4 years ago) (Cook et al., 1999; Zong et al., 1999).

The K15 gene, located at the right end of the genome, is at a position equivalent to the gene encoding LMP2A of EBV. It is a class II gene that occurs as two highly diverged alleles, P (predominant or prototype) and M (minor) with only 29% amino acid identity (Choi et al., 2000; Glenn et al., 1999; Poole et al., 1999). The K15 divergence resembles that seen in EBNA-2, EBNA-3A, EBNA-3B and EBNA-3C genes of EBV, each of which also occurs in two main forms but with relatively less divergence (Dambaugh et al., 1984; Sample et al., 1990). Thus EBV occurs as two main types, types 1 and 2. Each K15 allele comprises eight exons specifying a protein with 12 membrane-spanning domains and a C-terminal cytoplasmic domain. This structure resembles that of EBV LMP2A, except that LMP2A possesses an N-terminal, rather than a C-terminal, cytoplasmic domain (Sample et al., 1989). The cytoplasmic domain of K15 contains several YXXI/L motifs (reminiscent of LMP2A) and a putative TRAF binding site (as in LMP1). Transfection experiments showed that the K15 protein

(P allele) primarily localizes to plasma and cytoplasmic membranes (Choi et al., 2000; Glenn et al., 1999), and a CD8-K15 chimera stimulated with an anti-CD8 antibody significantly inhibited BCR signalling (Choi et al., 2000). Thus, K15 proteins possibly play a role in regulation of signal transduction (Poole et al., 1999).

PCR analysis of HHV-8 strains from various parts of the world indicates that the P allele is present in the great majority of HHV-8 genomes and is found in association with A, B, C and D K1 subtypes from all parts of the world (Lacoste et al., 2000a,b, Meng et al., 2001; Poole et al., 1999). The M allele, on the other hand, is rare and has thus far been found in association with the A, B and C subtypes from various parts of the world, including parts of Africa but not from East Africa (Lacoste et al., 2000a,b; Meng et al., 2001; Poole et al., 1999). The K15 allele linked to subtype E is unknown.

The origin of the K15 divergence is not clear, but one of the two K15 alleles may have resulted from a recombination event with a closely related γ_2-herpesvirus (Glenn et al., 1999; Poole et al., 1999).

The patterns of sequence variation observed in the internal (conserved) gene loci of various HHV-8 strains generally correlate with the K1 patterns (Poole et al., 1999). However, the K15 alleles appear to be essentially unlinked to the K1 patterns. Evidence for recombination (based on lack of co-segregation of strains at various loci) was reported in 20-30% of the HHV-8 strains analysed by Poole et al. (1999). Almost half (5 of 12) of the African strains analysed in this study displayed mosaic genomes that probably reflect a history of recombination within and among the major subtypes.

1.8. EPIDEMIOLOGY OF HHV-8

Detection of HHV-8 by both serology and PCR-based methods has been used to study the epidemiology and transmission of HHV-8. Accurate seroprevalence rates of HHV-8 remain unclear owing to the use of different tests and target antigens (Rabkin et al., 1998). However, the results point to the same conclusion. Unlike most herpesviruses, HHV-8 does not appear to be ubiquitous in most populations. The highest seroprevalence rates (>50%) have been reported for central and southern Africa, followed by the Mediterranean and eastern European countries (5-20%). The lowest prevalence (0-5%) occurs in North America, Northern Europe, most of Asia and the Caribbean countries (Table 1.4; reviewed by Chatlynne and Ablashi, 1999; Moore, 2000; Schulz, 1998). This pattern of infection matches the geographic distribution of KS (Table 1.4; Moore, 2000). The rates of AIDS-associated KS are highly dependent on local rates of infection with HIV-1, but not HIV-2 (Ariyoshi et al., 1998). The combined effects of widespread HIV-1 and HHV-8 infection have made KS the most common tumour in several sub-Saharan countries, including Uganda.

1.8.1. Seroprevalence of HHV-8

Serological assays have been developed that detect antibodies to both lytic and latent HHV-8 antigens, including the immunofluorescent (IF) test, Western blot (WB) analysis, radioimmunoprecipitation assay (RIPA) and ELISA (Table 1.5). The predominant latent antigen in PEL cell lines is LANA, which can be detected by WB (as a 224-236 kDa protein) or IF of untreated PEL cell lines (Gao et al., 1996b; Kedes et al., 1996; Simpson et al., 1996). By IF, LANA-positive cells show a characteristic nuclear stippling or speckled pattern. Antibodies to LANA are detected in 80-85% of patients with AIDS-KS, >90% of those with classic KS and only 0-3% of USA or UK blood donors (Gao et al.,

Table 1.4. Patterns of HHV-8 infection and KS among HIV-negative individuals

Regions	KS incidence	HHV-8 seroprevalence (%)	Route of transmission	Groups at risk
Northern Europe, Asia, North America	Low	0-5	Sexual, iatrogenic	Homosexual men, people with sexually transmitted diseases, transplant recipients
Mediterranean, Middle East	Intermediate	5-20	Sexual, iatrogenic, possibly non-sexual	Homosexual men, people with sexually transmitted diseases, transplant recipients, older adults
Africa, parts of Amazon basin	High	>50	Non-sexual, sexual	Children, young and old adults

Modified from Moore (2000).

Table 1.5. Serological assays used to detect HHV-8 antibodies in different cohorts

Antigen	Assay[a]	Antibody detection rate (%) in: Classic KS	AIDS KS	Homosexual men	Haemophilia	Blood donors	Other control groups	Reference
Latent antigen								
LANA (latency-associated nuclear antigen)	IF	85-94	71-88	18-30	0-3 (UK/USA)	0-3 (UK/USA)	12-21 (Greece)	Gao et al. (1996a,b) Kedes et al. (1996)
						4-21 (Italy)	51 (Uganda)	Simpson et al. (1996) Calabrò et al. (1998)
LNA (226/234 kDa nuclear protein)	WB	100	80-89	18-60		0 (USA)	62 (Uganda) 7 (USA)	Lennette et al. (1996) Gao et al. (1996a) Zhu et al. (1999)
Structural ('lytic') antigens								
40 kDa protein	WB		67	13				Miller et al. (1996)
ORF65	ELISA/WB	86-94	75-84		2 (UK)	1.7-5 (UK/USA) 9-22 (Italy)	12 (Greece) 35 (Uganda)	Simpson et al. (1996) Lin et al. (1997) Calabrò et al. (1998)
K8.1	WB	100	89	60		4	0 (Germany) 7 (USA)	Raab et al. (1998) Zhu et al. (1999)
ORF26	ELISA		~40 ~60			~11 (Germany) ~20 (USA)		André et al. (1997) Davis et al. (1997)
Unidentified	IF		96-100	90-100		20-25 (USA)		Lennette et al. (1996)
	ELISA/IF		100			0-12 (USA)	32-100 (Africa)	Chatlynne and Ablashi (1997) Whitman et al. (1997)
	IF		100			0 (USA)		Smith et al. (1997)
Mixed-antigen (K8.1, ORF59, ORF65, ORF73)	ELISA		100				1.4-1.9	Katano et al. (2000a)

[a] IF, immunofluorescence test; WB, Western blot.

Modified from Schulz (1998).

1996a,b; Kedes et al., 1996; Lennette et al., 1996; Simpson et al., 1996). LANA has no homologue in EBV (Russo et al., 1996), thus eliminating the possibility of antigenic crossreactivity with this highly prevalent γ-herpesvirus.

To address the possibility that LANA assays may underestimate HHV-8 prevalence, several groups have developed assays based on lytic (structural) viral proteins. The minor capsid protein encoded by ORF65 and the K8.1 glycoprotein have been identified as potent and useful antigens for serological analysis (Chandran et al., 1998a,b; Li et al., 1999; Lin et al., 1997; Pau et al., 1998; Raab et al., 1998). Antibodies to these and other proteins have been detected by IF test, WB and RIPA in chemically induced PEL cell lines (Table 1.5). Whole HHV-8 virions purified from induced PEL cell lines have also been used successfully in ELISA. However, the specificity of the assays based on lytic proteins is debatable due to the high homology of certain HHV-8 structural proteins with their EBV counterparts (André et al., 1997; Moore et al., 1996b). Recently, a mixed-antigen ELISA containing four proteins (K8.1, ORF59, ORF65 and ORF73) detected 100% HHV-8 seropositivity in sera from Japanese KS patients and only 1.4-1.9% in control groups (Table 1.5). No cross-reactivity with HCMV or EBV was detected in this assay, indicating that it is specific. ORF65 ELISA detects antibodies in 75-80% of AIDS-KS, 85-90% of classic KS, 1-3% of UK blood donors and 5-10% of USA blood donors. The assay therefore has sensitivity and specificity comparable to that of LANA IF and WB tests. Recombinant K8.1 has also been shown to have high sensitivity and specificity (Table 1.5). However, ORF26 (major capsid protein)-based assays detect antibodies in approximately 40-60% of KS sera and 20% of blood donor sera (André et al., 1997; Davis et al., 1997). Antibodies to ORF26 protein cross-react with EBV (André et al., 1997).

Among UK and USA blood donors, serological reactivity to LANA, ORF65 and unidentified structural antigens ranges from 0-3%, 1-5%, and 0-25%, respectively (Table 1.5). In Italy and Greece, reactivity in blood donors ranges from 4-25%. The relatively high HHV-8 prevalence in Italy is broadly in agreement with the higher incidence of classic KS in this region. In Africa, antibodies to LANA have been found (with considerable regional variation) in a relatively higher proportion (6-53%) of blood donors (Ariyoshi et al., 1998; Gao et al., 1996b; Lennette et al., 1996; Simpson et al., 1996). While HHV-8 appears to be widespread in Africa, endemic KS is restricted to East and Central Africa (Beral, 1991; Gompels and Kasolo, 1996), suggesting differences in HHV-8 strain pathogenicity or involvement of other co-factors. For example, in The Gambia (West Africa) where HHV-8 seroprevalence is approximately 60% in antenatal HIV-negative mothers, endemic KS is rare and AIDS-KS is virtually limited to individuals infected with HIV-1 but not with HIV-2 (Ariyoshi et al., 1998). This supports a co-factor role for HIV-1 (Ensoli et al., 1994; Albini et al., 1996).

Among HIV-infected individuals from the USA and Northern Europe, HHV-8 antibodies are found significantly more frequently in HIV-positive homosexual men (30-35%) than in other HIV risk groups, such as haemophiliacs and intravenous drug users (Gao et al., 1996b; Kedes et al., 1996; Lennette et al., 1996; Simpson et al., 1996). This suggests sexual transmission of the etiologic agent. Furthermore, the distribution of HHV-8 in HIV risk groups correlates with the reported high incidence of KS in HIV-infected homosexual men and the rarity of KS in HIV-infected patients with haemophilia or intravenous drug users (Beral, 1991; Beral et al., 1990; Peterman et al., 1993), thus providing an argument for HHV-8 being the cause of KS. The seroprevalence of HHV-8 among HIV-infected women (4%) in the USA is much lower than that in HIV-infected homosexual men (30-35%) (Gao et al., 1996a,b; Kedes et al., 1996;

Simpson et al., 1996). This, again, is consistent with the reported lower incidence of KS among women than in HIV-infected men, further extending the argument for the causative role of HHV-8. Longitudinal serologic studies of HIV-positive men show that seroconversion precedes KS development (Gao et al., 1996a,b). Thus, in contrast to HHV-8 endemic areas, presence of HHV-8 antibodies in individuals in non-endemic areas (USA and Northern Europe) is strongly associated with KS. HHV-8 seroprevalence among HIV-infected individuals in endemic areas ranges from 36-91%, with higher prevalence rates in individuals with KS (reviewed by Chatlynne and Ablashi, 1999).

1.8.2. Molecular epidemiology

PCR-based methods have been used to detect HHV-8 DNA in KS lesions, PBMCs, lymphoid tissue, saliva, and semen of KS patients and healthy individuals. The HHV-8 genome has been detected in virtually all KS lesions (Boshoff et al., 1995b; Buonaguro et al., 1996; Chang et al, 1994, 1996b; Chuck et al., 1996; Dictor et al., 1996; Dupin et al., 1995; Gaidano et al., 1996b; Luppi et al., 1996b; Moore and Chang, 1995; Noel et al., 1996; Schalling et al., 1995; Su et al., 1995). HHV-8 DNA was also detected in 14% (3/22) of non-KS tumour lesions from Ugandan control cancer patients (Chang et al., 1996b). HHV-8 DNA can be detected in 50-60% of PBMC samples from KS patients (Whitby et al., 1995). Thus, this method underestimates HHV-8 prevalence. HHV-8 is not, or only infrequently, detected in PBMCs from HIV-negative individuals (Bigoni et al., 1996; Brambilla et al., 1996; Humphrey et al., 1996; Lefrere et al., 1996; Moore et al., 1996c; Whitby et al., 1995). It has been detected in 10% of PBMCs from HIV-negative individuals in The Gambia (Ariyoshi et al., 1998) and in 0% (0/21) of PBMCs from KS-negative individuals in Uganda (Purvis et al., 1997). Detection of HHV-8 DNA in PBMCs occurs at a relatively higher frequency in HIV-infected individuals and its presence predicts

progression to KS (Whitby et al., 1995). The HHV-8 genome has been detected in the lymphoid tissue of about 10% of healthy Italian subjects (Bigoni et al., 1996; Viviano et al., 1997).

Using quantitative PCR analysis in a longitudinal study involving HHV-8-positive, KS-negative men from the USA who have sex with men, Pauk et al. (2000) detected HHV-8 in 30% of oropharyngeal samples, as compared to 1% of anal and genital samples. HHV-8 was detected in saliva in 39% of these men on more than 35% of the consecutive days on which samples were obtained. HHV-8 has also been detected in the saliva of KS patients by some studies (Boldough et al., 1996; Koelle et al., 1997) but not by others (Ambroziak et al., 1995; Whitby et al., 1995), and in the semen of KS patients (Corbellino et al., 1996a,c; Gupta et al., 1996; Howard et al., 1997; Monini et al., 1996a,b). Although the virus is generally not detected in the semen of healthy men (Corbellino et al., 1996a,c; Gupta et al., 1996; Howard et al., 1997; Lebbé et al., 1997; Lefrere et al., 1996; Moore et al., 1996c; Whitby et al., 1995), it has been detected in 13-23% of semen samples from healthy Italian donors from areas where classic KS appears more frequently (Blackbourn and Levy, 1997; Monini et al., 1996a,b; Staskus et al., 1997; Viviano et al., 1997).

1.8.3. Transmission of HHV-8

Little is known about the precise way in which HHV-8 is transmitted. There is evidence from several studies to suggest that it can be sexually transmitted, and this may be the major route of transmission in non-endemic areas. HHV-8 antibodies are more frequently found in sexually transmitted diseases (STD) clinic attendees (Kedes et al., 1996, 1997; Simpson et al., 1996; Lennette et al., 1996). In the USA, HHV-8 seroprevalence among homosexual men is strongly associated with promiscuity and a history of STDs (Martin et al., 1998). Also

among Danish homosexual men, HHV-8 seropositivity and seroconversion are associated with promiscuity, duration of homosexual activity, receptive anal intercourse and, in the early 1980s, sexual contact with homosexual men from the USA (Melbye et al., 1998). These behavioural risk factors have been shown previously to increase the risk of KS developing in homosexual men (Beral et al., 1990), thus supporting the notion that HHV-8 plays a causative role in the development of KS. Additional evidence that HHV-8 may be primarily sexually transmitted in non-endemic countries comes from data indicating that HHV-8 infection is rare in infants and children, but occurs mainly after puberty in these areas (Blauvelt et al., 1997; Lennette et al., 1996; Raab et al., 1998; Regamey et al., 1998). Evidence for possible heterosexual transmission has also been documented. Sitas et al. (1999) reported a statistically significant association between the number of heterosexual partners and HHV-8 seropositivity among black cancer patients in South Africa. Furthermore, a large study by Davidovici et al. (2001) of HHV-8 seroprevalence among Jewish families in Israel indicated that HHV-8 can be transmitted between heterosexual married couples. Semen could be a source of infection in sexual transmission. However, the study by Pauk et al. (2000) suggests that saliva rather than semen or anal secretions may play a more significant role in the transmission of HHV-8 in men who have sex with men. Indeed, one study has showed that HHV-8 from saliva of men with KS could be transiently propagated in 293 cells, indicating that it was infectious (Vieira et al., 1997).

Factors favouring HHV-8 transmission in endemic African countries (where HHV-8 infection is widespread in both men and women) appear to be different from those in non-endemic countries. Antibodies, steadily increasing with age, have been detected to HHV-8 LANA and ORF65 proteins in young children (aged 2-12 years) in East Africa and Italy (Calabrò et al., 1998; Mayama et al., 1998). These results suggest that horizontal, non-sexual, routes of transmission

may be more important in areas where KS is more common. HHV-8 has also been detected by PCR in PBMCs of 8% of young children in Zambia (Kasolo et al., 1997). In the study by Mayama et al. (1998), which involved Ugandan children and adolescents, the presence of HHV-8 antibodies was independently associated with hepatitis B virus (HBV) infection. This finding implies that both HHV-8 and HBV are transmitted via similar routes in this population or that factors associated with HBV infection also favour HHV-8 transmission.

The mode of transmission in children is not known, but saliva, through some form of casual contact, may be responsible for spreading HHV-8 in endemic areas. One study of a Sardinian population (where KS is highly endemic) found clustering of HHV-8 seropositivity among family members, suggesting possible vertical or, more likely, horizontal transmission within a family (Angeloni et al., 1998). The study by Mayama et al. (1998) showed that of 9 children younger than 2 years, only 1 infant (6 months old) had antibodies to HHV-8. This finding does not exclude vertical transmission from mother to child, but overall results of this study suggested that this might not be the predominant mode of HHV-8 infection among young children in this cohort. Among Jewish families in Israel, Davidovici et al. (2001) found that for a child to test positive, the most important risk factor was maternal seropositivity, suggesting that mother to child transmission was important. All the children in this study were older than 2 years; thus a positive test indicated exposure to the virus since transplacental (passive) maternal antibodies would not be expected to be present at this age. Transmission of HHV-8 from sibling to sibling (in addition to from mother to child) has also been shown to be important in an African population living in French Guiana (Plancoulaine et al., 2000).

Iatrogenic transmission of HHV-8 is also of concern. Transmission through solid organ allografts has been documented (Luppi et al., 2000; Parravicini et al.,

1997b). However, it is apparent that most transplantation-associated KS is due to HHV-8 reactivation (Parravicini et al., 1997b). Blood transfusion does not appear to present a particular risk in HHV-8 transmission (Beral et al., 1990; Calabrò et al., 1998; Whitby et al., 1998), although a single case of infectious HHV-8 from a healthy blood donor has been documented (Blackbourn et al., 1997).

1.9. CONCLUDING REMARKS

Since the discovery of HHV-8 in 1994, a strong case has developed for its causative role in KS. Epidemiologically, HHV-8 is found in all forms of the disease, irrespective of HIV status, and infection tracks closely with known risk groups for KS in non-endemic areas. Furthermore, infection precedes the development of the tumour and is predictive of increased probability of KS development in HIV-infected individuals. In countries, and regions within countries, where KS occurs more frequently, HHV-8 is also more common; however, the converse is not always true. Thus, although HHV-8 appears to be required for development of KS, it is evident that other factors (e.g. other infectious agents, like HIV-1, immunosuppression or environmental factors) are necessary. Biologically, HHV-8 is closely related to other oncogenic rhadinoviruses and encodes several proteins that have been shown (in vitro) to have signalling properties and to control cellular growth. Furthermore, HHV-8 infection is targeted to the cell type (spindle cell) thought to play a key role in the pathogenesis of KS. Finally, some herpesvirus DNA polymerase inhibitors have been shown to reduce the appearance of KS lesions in AIDS patients. Taken together, these data strongly suggest that HHV-8 is a cause of KS.

1.10. JUSTIFICATION AND OBJECTIVES OF THE STUDY

1.10.1. Justification of the study

KS is a prominent disease in Uganda accounting for half the reported cancer cases (Wabinga et al., 1993). It occurs in epidemic and endemic forms affecting both children and adults. At the time this study was initiated, little was known about the prevalence of HHV-8 in the Ugandan population or the strains present in this country. I performed a study to determine the prevalence of HHV-8 in Ugandan blood donors and to characterise the HHV-8 strains in Ugandan KS patients.

1.10.2. Specific objectives of the study

1. Determination of HHV-8 prevalence in HIV-negative blood donors.
2. Characterization of the HHV-8 K1 gene in tumour DNA from KS patients.
3. Characterization of the HHV-8 K15 gene in tumour DNA from KS patients.
4. Characterization of selected HHV-8 internal gene loci in tumour DNA from KS patients.

CHAPTER 2
MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Blood donors

One hundred and sixteen blood donors visiting Nakasero blood bank (NBB) in Kampala, the capital city of Uganda, during April-August 1998 were recruited into the study. NBB is the only blood bank facility serving the central Uganda region.

2.1.2 Kaposi's sarcoma patients

Twenty eight unrelated, adult Ugandan KS patients attending the Uganda Cancer Institute (UCI) clinic at Mulago Hospital, Kampala, during April-July 1998 were recruited to the study. Samples had also been collected from two other patients (Ugd1 and 2) in August 1997 bringing the total number of KS patients to 30. The patients were at various stages of anticancer chemotherapeutic treatment. Mulago Hospital is the major national referral hospital, and UCI receives KS referrals from across the country. UCI has a 40-bed unit for admitted patients.

Skin biopsy punches (6 mm in diameter) used to collect skin biopsies from KS patients were obtained from Rocialle Medical Limited, Sawston, Cambridge, UK.

2.1.3 DNA extraction and Polymerase Chain Reaction (PCR)

Nucleon BACC 2/3 for cell pellet DNA extraction	Nucleon Biosciences
Taq DNA polymerase, 10X PCR buffer with magnesium chloride	Boehringer-Mannheim or Sigma
Primers	Initially synthesized at the Institute of Virology and later obtained from MWG-Biotech AG
dNTPs mixture (each 12.5 mM)	Amersham Life Sciences
Mineral oil	Sigma
Aerosol resistant tips (ART)	Gibco Life Technologies

2.1.4 Electrophoresis

High and low melting point agarose and ethidium bromide (10 mg/ml aqueous solution)	Sigma
DNA markers (lambda DNA *Hin*d III digest, 1 kbp, 100 bp and 123 bp ladders)	New England BioLabs

2.1.5 Purification, cloning and sequencing of DNA fragments

Hybaid DNA purification kit	Hybaid
pGEM-T Vector System II (including JM 109 competent cells)	Promega
Restriction endonucleases and other enzymes and buffers	Boehringer-Mannheim, BRL or New England Biolabs
ABI PRISM Termination Ready Reaction Mix	Perkin-Elmer
Internal gene primers for sequencing	Synthesized at the Institute of Virology and later supplied by MWG-Biotech AG
M13 Universal Sequencing Primers	Pharmacia Biotech

2.1.6 Chemicals

All chemicals were obtained from BDH, Boehringer-Mannheim or Sigma.

2.1.7 Solutions and buffers

TE	10 mM Tris-HCl, pH 8 1 mM EDTA

Extraction buffer	10 mM Tris-HCl, pH 7.5 5 mM EDTA 0.5% SDS 0.5 mg/ml proteinase or 100 μg/ml proteinase K
10X TBE	109 g/l Tris 55 g/l boric acid 9.3 g/l EDTA
DF dyes	37.2 g/l EDTA 100 g/l Ficoll 400 5X TBE 1% (w/v) bromophenol blue
5X ligase buffer	250 mM Tris-HCl, pH 7.5 50 mM $MgCl_2$ 5 mM DTT 25% (w/v) PEG 6000
X-gal	40 mg/ml 5-bromo-4-chloro-3-indoyl β-D-galactoside in N,N'-dimethyl formamide
IPTG	30 mg/ml isopropylthio-β-D-galactoside

PEG/NaCl	20% (w/v) PEG 6000 2.5 M NaCl
Phenol TE	phenol equilibrated with TE

2.1.8 Bacterial growth media

2YT broth	85 mM NaCl 1% (w/v) bactopeptone 1% (w/v) yeast extract
L-Broth	177 mM NaCl 1% (w/v) bactopeptone
L-Broth Agar	1.5% (w/v) agar in L-broth
Top agar	0.6% (w/v) bacto-agar in water

2.1.9 Published sequences

Published sequences used in this study were obtained from the following references: Alagiozoglou et al. (2000); Cook et al. (1999); Glenn et al. (1999); Lacoste et al. (2000a); Nicholas et al. (1998); Poole et al. (1999); Russo et al. (1996); Zong et al (1999). The sources of some are listed in Table 2.1.

Table 2.1. Sources of some of the published sequences used in this study

Strain	K1 subtype	K15 allele	Source	Country	Reference (GenBank accession)
BC-1	A2	M	AIDS-PEL cell line	USA	Russo et al. (1996) (U75689)
BCBL-R	A1	P	AIDS-PEL tumour cells	USA	Nicholas et al. (1998) (U85269)
ASM72	C1	M	Autopsy: AIDS disseminated KS	USA	Zong et al. (1999)
BC2	C3	P	AIDS-PEL cell line	USA	Zong et al. (1999
GK18	C	P	Classical KS biopsy	Greece	Glenn et al. (1999) (AF148805)
K1-43/Ber	Undefined	M	HIV positive, PEL ascitic fluid	France	Lacoste et al. (2000a) (AF178810)
ST1	B	Unknown	AIDS-KS biopsy	Uganda	Zong et al. (1999)
ST2	B	Unknown	AIDS-KS biopsy	Uganda	Zong et al. (1999)
Ug374	A5	Unknown	HIV positive, no KS, PBMC	Uganda	Cook et al. (1999) (AF130289)
431KAP	B	P	Endemic KS biopsy	Zaire	Zong et al. (1999)
OKS3	A5	P	AIDS-KS biopsy	Tanzania	Zong et al. (1999)
OKS4	B	P	AIDS-KS biopsy	Tanzania	Zong et al. (1999)
K1-8/Dem	Undefined	No product	AIDS-KS biopsy	Central African Republic	Lacoste et al. (2000a) (AF178780)
DS814ZA	Unknown	Unknown	HIV positive, no KS, PBMC	South Africa	Alagiozoglou et al. (2000) (AF243803)
KS70ZA	Unknown	Unknown	AIDS-KS, PBMC	South Africa	Alagiozoglou et al., (2000) (AF243831)
KS84ZA	Unknown	Unknown	AIDS-KS, PBMC	South Africa	Alagiozoglou et al. (2000) (AF243823)
KS91ZA	Unknown	Unknown	AIDS-KS, PBMC	South Africa	Alagiozoglou et al. (2000) (AF243823)
TKS10	D1	P	Classic KS	Taiwan	Zong et al. (1999)

2.2 METHODS

2.2.1 Blood donors

Individuals donating blood during the times I was at NBB were sampled. As well as collection of blood samples, demographic data, including age, gender, place of residence and birth were recorded. Blood samples were collected in 7 ml EDTA-tubes. Plasma was harvested and stored at -70 °C at the Makerere University-Johns Hopkins University (MU-JHU) collaborative laboratory, Mulago Hospital, until shipped on dry ice to the UK. A cell pellet (containing all white blood cells) was prepared from the rest of the sample by staff at the MU-JHU laboratory using the method involving lysis of red blood cells.

All samples were tested for HIV-1 antibodies using two commercially available ELISA kits (Burroughs Wellcome and Welcozyme).

2.2.2 Kaposi's sarcoma patients

The HHV-8-positive DNA samples used for genome characterization were derived from skin biopsies of KS patients. Ethical approval for this study was obtained from the Uganda AIDS Research Committee and the Uganda National Council for Science and Technology. Informed consent was obtained from each patient before samples were taken. Samples were taken from randomly selected in-patients meeting the following inclusion criteria: >18 years old (except one patient, Ugd19, who was 16 years old), having clinically typical KS with haematological indices acceptable for therapy (Hb>10 g/dl, WBC >3.0 X 10^9/l with absolute neutrophil count of >1.5 X 10^9/l and platelets >100 X 10^9/l) and a signed consent form. Patients with a bleeding tendency or with small lesions, and pregnant women in the first trimester were excluded. All patients, except Ugd30,

were clinically and histologically diagnosed by Dr. Edward Katongole-Mbidde (UCI). Ugd30 became unavailable after preliminary clinical diagnosis. The patients were tested for HIV-1 antibodies at the MU-JHU collaborative laboratory using two commercial ELISA kits (Burroughs Wellcome and Welcozyme). Samples that reacted non-specifically were confirmed by Western blot (WB) analysis.

The biopsies were collected by Sister Naomi Byabazaire. She immediately dropped them into absolute ethanol kept on liquid nitrogen to minimize DNA degradation. The ethanol was removed following overnight incubation at -80 °C, and the biopsies were stored at 4° C until extracted. Matching blood samples were collected in EDTA-tubes for HIV-1 testing, HHV-8 serology and for preparation of cell pellets as described above.

2.2.3 Serological tests

Serological tests were performed on all blood donor samples and some (19) KS patient plasma samples by Julie Sheldon in the laboratory of Professor Thomas Schulz (Department of Medical Microbiology and Genitourinary Medicine, University of Liverpool, Liverpool, UK). Diluted (1:100) plasma were screened for antibodies to HHV-8 ORF65 and ORF73 recombinant proteins by ELISA, as described previously (Simpson et al., 1996). The antigens were diluted 1:60 (ORF65 and control) or 1:50 (ORF73) in 100 mM $NaHCO_3$ (pH 8.5). The plates were coated at room temperature overnight. The cut-off values for the ELISA scores, (+), 1+, 2+ and 3+, were, respectively, 3, 5, 7 and 10 standard deviations above the average value for 10 negative control sera. An immunofluorescence (IF) test for LANA (Rainbow et al., 1997) was then performed on all samples on paraformaldehyde-fixed BCP-1 cells (Boshoff et al., 1998) at a plasma dilution of 1:50, as described previously (Simpson et al., 1996). WB analysis against

ORF65 protein was performed for all samples that were positive in the ELISA (one or both antigens) but negative or non-specific in the IF test. The results of the IF test and the WB analysis were scored arbitrarily on the basis of the intensity of the signal. Overall, samples were interpreted as being positive if they were positive in both the IF test and the ELISA (one or both antigens), or in the IF test alone or in the ELISA alone followed by a positive WB result.

2.2.4 DNA extraction

Tumour DNA was extracted from biopsies and cell pellet samples at the MU-JHU laboratory. Biopsies were extracted as described previously for fresh frozen biopsy samples (Cook et al., 1999). Briefly, ethanol-treated tissue was cut into small pieces with a scalpel and incubated overnight at 55 °C in extraction buffer. Proteinase K was inactivated by incubating at 95 °C for 15 min. DNA was extracted using phenol and chloroform-isoamyl alcohol, and precipitated with 2.5 volumes of 100% ethanol and 0.1 volumes of 3M sodium acetate (pH 5.5) at room temperature for 5 min. The pellet was then rinsed twice with 70% ethanol. Cell pellet DNA was extracted using a commercial kit (BACC 2/3; Nucleon Biosciences) as instructed in the manufacturer's manual. Precipitated DNA was resuspended in 100-300 µl TE and stored at 4 °C until shipped at ambient temperature to the Institute of Virology, Glasgow, UK.

2.2.5 PCR and DNA sequencing

Tumour DNA was amplified by PCR for seven loci of the HHV-8 genome: K1 (1277 bp), K3 (635 bp), ORF26 (160, 233, 571 bp), K9 (594 bp), T0.7/K12 (648 bp), ORF75 (749 bp for genomes with the K15 P allele and 1487 bp for the genome (Ugd10) with the K15 M allele), K15 P (285 bp to identify the allele and 2494 bp to sequence it completely) and K15 M (298 bp to identify the allele and

two overlapping regions of 1374 and 1483 bp to sequence it completely). The locations of the ORF and the PCR products are shown in Table 2.2. The sizes of the PCR products and corresponding primer pairs are listed in Table 2.3. Primary PCR products (i.e. not nested) were obtained, except for the K1 locus in three samples (Ugd4, Ugd12 and Ugd24) which was obtained as a 874 bp product by nested PCR and for ORF26 in the detection of HHV-8 DNA in cell pellet samples. As a control the human β-globin gene was amplified using primers shown in Table 2.3.

PCR was performed using *Taq* polymerase in a total volume of 50 µl containing 20-200 ng of template DNA. Stringent measures were undertaken to prevent PCR contamination: all original DNA samples were extracted in a laboratory in Uganda; separate rooms were used for pre-PCR, PCR and post-PCR steps; aerosol resistant tips were used and the premix was UV-irradiated. Each experiment involved a sizeable batch of reactions, and incorporated appropriate negative (water and uninfected cellular DNA) and positive controls. The conditions used to amplify the products are shown in Table 2.3.

PCR products were purified directly from solution or from 1% low melting point agarose gels stained with ethidium bromide using a commercial kit (Hybaid). K1, K15 and ORF75 (linked to K15 M allele) products were cloned into pGEM-T. Plasmid (template) DNA for sequencing was prepared by the alkaline lysis method followed by PEG/NaCl precipitation. Sequencing was performed using universal and gene-specific primers except for some samples (K15 gene of Ugd2 and initial clones of ORF75 and K15 products of Ugd10), where the shotgun M13 sequencing technique was employed. The K15 product of Ugd2 was sequenced by Andrew Davison and Charles Cunningham. The initial, almost complete, sequence of the ORF75 and K15 products of Ugd10 was used to design internal primers for sequencing subsequent clones. Three independent

Table 2.2. Locations[a] of the genes, PCR products and sequences analysed in this and previous studies

Gene	ORF	PCR product	Region analysed in this study	Region analysed/sequenced in previous study	Reference to previous study
K1	105-974	68-1344 (primary) 101-974 (nested)	120-959	120-959	Cook et al. (1999) Zong et al. (1999)
K3	19609-18608	18965-19599	19052-19570	none	
ORF26	46933-47850	47006-47576	47014-47508	47193-47522	Poole et al. (1999)
K9	85209-83860	83900-84493	83954-84453	none	
T0.7/K12	118101-117919	117469-118116	117518-118065	117469-118114	Poole et al. (1999)
ORF75	134441-130550	133008-134494 (M) 133824-134572 (P) (*726-1474*)	133864-134441 (ORF75) *(766-1343) (P)* 134442-134650 (UPS75') (*1344-1552*) [18603-19389]	132875-133728 (ORF75-E) 133639-134442 *631-2130* (UPS75) [1-28559]	Poole et al. (1999) Alagiozoglou et al. (2000) Zong et al. (1997) Glenn et al. (1999)
K15 M	136772-134672	135607-136980 (I) 134422-135904 (II)	134442-136980	1-137508	Russo et al. (1996)
K15 P	*3652-1578* [21490-19415]	*1369-3862* [19161-21700]	*1369-3862* [19161-21700]	*1-4623* [1-28559]	Nicholas et al. (1998) Glenn et al. (1999)

[a] Coordinates are from the BC-1 genome sequence (Russo et al., 1996), except for those in italics which are from the portion of the BCBL-R sequence published by Nicholas et al. (1998) or those in square brackets which are from the portion of GK18 sequence published by Glenn et al. (1999).

Table 2.3. PCR primer pairs and products

Gene locus[a]	Product size (bp)	Primer	Sequence	Location in BC-1/BCBL-R[b]	PCR conditions
K1[c] (FL)	1277	PMC20	GTCTTTCAGACCTTGTTGG	68-86	94 °C for 45 s, 47.5 °C for 30 s, 72 °C for 1 min over 30 cycles
		PMC21	CATGTTGCTGTTATATTGCG	1344-1325	
K1[c] (FL)	874	O1	CCTGTGGATCCAAGATGTTCCTGTATG	101-117	94 °C for 45 s, 57 °C for 30 s, 72 °C for 1 min over 30 cycles
		O2	ACCTGGAATTCAGTACCAATCCACTG	974-958	
K3	635	K3-A	AGGATGTTCCTGTCTGCTGGA	19599-19579	94 °C for 45 s, 60 °C for 30 s, 72 °C for 1 min over 30 cycles
		K3-B	ATCTTGGTCGCCTGGAGCTGC	18965-18985	
ORF26[c]	571	KS5	GACTCTTCGCTGATGAACTGG	47006-47026	Same as for K3
		KS4	AGCACTCGCAGGGCAGTACG	47576-47557	
	233	KS1	AGCCGAAAGGATTCCACCAT	47287-47307	94 °C for 45 s, 55 °C for 30 s, 72 °C for 45 s over 25 cycles
		KS2	TCCGTGTTGTCTACGTCCAG	47519-47499	
	160	NS1	ACGGATTTGACCCCGTGTTC	47321-47341	94 °C for 2 min over 1 cycle; 94 °C for 1 min, 65 °C for 1 min, 72 °C for 1 min over 45 cycles
		NS2	AATGACACATTGGTGGTATA	47480-47460	
K9	594	K9-D1	GTGCCATCTTGTACGACGGCA	84493-84473	Same as for K3
		K9-D2	GTACGACGCAGGCGTCTGAGA	83900-83920	
T0.7/K12[c]	648	LGH2075	CTCCAATCCCAATGCATGGA	118116-118097	Same as for K3
		LGH2076	GCTGCAATGTACTGCCATG	117469-117487	
ORF75 M	1487	K15-8A	TCTCGGCAGCCTGACTACAG	134494-134476	Same as for K1 (1277 bp product)
		K15-9A	CGACTCTGCCACCACCGCA	133008-133026	
ORF75 P	749	K15-24C	ACTCTGCAGAGCAACCTGTGC	*1474-1454*	Same as for K3
		K15-25C	TGAAGCCTGCACCGCCAGCG	*726-745*	
K15 M	298	K15-3A	TCCCATTCATGCCTTTCTGT	135904-135885	Same as for K1 (1277 bp product)
		K15-4A	CTTCGTTATTGGTGTCTTGT	135607-135626	
K15 M (I)	1374	K15-6A	CACCAGGATGCAGTGTTACA	136980-136961	Same as for K1 (1277 bp product)
		K15-4A	CTTCGTTATTGGTGTCTTGT	135607-135626	

K15 M (II)	1483	K15-3A K15-2	TCCCATTCATGCCTTTCTGT CCAGTGACGTCGTAGGCCAT	135904-135885 134422-134441	Same as for K1 (1277 bp product)
K15 P	285	K15-3C K15-4C	ACGCATACATGTACTGCCAC CTTTGATATTGCCAGTGGTG	*2785-2766* *2501-2520*	Same as for K1 (1277 bp product)
K15 P (FL)	2494	K15-6C K15-7C	ACAATTTACGAGCCTTGTATC GACACCCTCTGTAGTCAGACT	*3862-3842* *1369-1389*	Same as for K1 (1277 bp product)
β-globin[d]	110	PCO_3 P64	ACACAACTGTGTTCACTAGC CAACTTCATCCACGTTCACC	- -	94 °C for 5 min over 1 cycle; 94 °C for 1 min, 55 °C for 1 min, 72 °C for 1 min over 35 cycles; 72 °C for 1 min over 5 cycle

[a] M, strain (Ugd10) containing the K15 M allele; P, strains containing the K15 P allele; FL, the full length gene was amplified.

[b] Coordinates from the BC-1 genome sequence (Russo et al., 1996) for most products, or from the portion of the BCBL-R genome sequence published by Nicholas et al. (1998) for the ORF75 and K15 loci in strains containing the K15 P allele (in italics). Those for β-globin are not given.

[c] Published primers (Bigoni et al., 1996 ; Boshoff et al., 1995; Chang et al., 1994; Cook et al., 1999; Poole et al., 1999).

[d] Primers provided by Ashraf Basuni.

clones were sequenced for each product to exclude PCR-induced errors. All other products were sequenced directly using the appropriate PCR primers. The sequencing reactions were performed with the ABI-PRISM Big Dye Terminator cycle sequencing ready reaction kit and the sequences were determined on an ABI PRISM 377 DNA sequencer. Data were derived from both DNA strands for all products.

2.2.6 DNA sequence analysis

DNA sequences were assembled and edited using SAP or Pregap4/Gap4 (Staden, 1987; Staden et al., 1998) and aligned using Pileup and Pretty (GCG, Madison, Wisconsin). K1 sequences were analysed phylogenetically together with published sequences (Cook et al., 1999; Lacoste et al., 2000a; Zong et al., 1999) by the neighbour joining method, using Seqboot, Dnadist, Neighbor and Consense from the PHYLIP package version 3.572 (University of Washington, Seattle). Phylogenetic analysis was performed with the help of Duncan McGeoch (MRC Virology Unit, Institute of Virology, Glasgow, UK). DNA sequences were translated into amino acid sequences using Translate (GCG), and divergence values between pairs of amino acid sequences were determined using Protdist (PHYLIP).

Network analysis (Bandelt et al., 1995) of DNA sequence data was used to determine genotypes at all other loci with extensive assistance from Rory Bowden (Department of Statistics, University of Oxford, Oxford, UK). Networks provide an objective summary of sequence data and an effective way to infer a likely history of a set of closely related sequences. They offer an advantage over phylogenetic trees in studying variation between very similar sequences, and facilitate identification of potential sequencing errors.

CHAPTER 3
RESULTS AND DISCUSSION-I
PREVALENCE OF HHV-8 IN UGANDAN BLOOD DONORS

3.1 INTRODUCTION

It is now evident from the literature that HHV-8 infection is not found uniformly throughout the world's population. Several studies have shown that HHV-8 seroprevalence in African populations is significantly higher than elsewhere in the world, and is not limited to adults or HIV-infected individuals (Calabrò et al., 1998; Gao et al., 1996; Mayama et al., 1998; Sarid et al., 1999; Simpson et al., 1996; Schulz, 1998). Several studies have determined the prevalence of HHV-8 antibodies among different groups of the Ugandan population using various serological tests. The results of these studies (summarized in Table 3.1) show that HHV-8 prevalence in individuals who are not patients and do not have KS or HIV ranges from 11-77%. The prevalence of HHV-8 in blood donors has been reported by one study which employed the whole virus ELISA method as being 38% (Chatlynne and Ablashi, 1999). The prevalence in HIV-infected individuals, with or without KS, ranges from 46-89%, and that in KS patients ranges from 78-100%. HHV-8 prevalence in children and adolescents in Uganda was found to be 43% and 33% by ORF65 ELISA (confirmed by Western blot) and LANA IF, respectively, and 52% by both tests combined.

Several groups have failed to detect HHV-8 by PCR in the peripheral blood of healthy donors (Ambroziak et al., 1995; Whitby et al., 1995), while others have reported detection in 9% of PBMC and lymphoid tissue of HIV-uninfected individuals (Bigoni et al., 1996). PCR and Southern blot analysis of PBMC from KS-negative, HIV-positive Ugandan blood donors indicated that 0 of 10 samples

Table 3.1. HHV-8 seroprevalence among different groups of the Ugandan population

Group	Percent positive/ number tested	Test	Reference
Individuals without HIV or KS			
HIV-negative without KS	51/47	BCP-1[a] IF	Gao et al., 1996
HIV-negative without KS	62/47	LANA WB	Gao et al., 1996
HIV-negative without KS	35/17	ORF65 ELISA	Simpson et al., 1996
HIV-negative without KS	53/17	Latent IF	Simpson et al., 1996
Blood donors	38/58	Whole virus ELISA	Chatlynne & Ablashi, 1999
General population	77/82	Lytic BCBL-1 IF	Lennette et al., 1996
General population	11/82	Latent BCBL-1 IF	Lennette et al., 1996
HIV and/or KS patients			
HIV-negative with KS	100/21	Lytic BCBL-1[a] IF	Lennette et al., 1996
HIV-negative with KS	100/21	Latent BCBL-1 IF	Lennette et al., 1996
AIDS-KS	78/18	BCP-1 IF	Gao et al., 1996
AIDS-KS	89/18	LANA WB	Gao et al., 1996
AIDS-KS	82/17	ORF65 ELISA	Simpson et al., 1996
HIV-positive without KS	51/35	BCP-1 IF	Gao et al., 1996
HIV-positive without KS	71/35	LANA WB	Gao et al., 1996
HIV-positive without KS	46/35	Whole virus ELISA	Chatlynne & Ablashi, 1999
HIV-positive without KS	47/34	ORF65 ELISA	Simpson et al., 1996
HIV-positive without KS	53/34	Latent IF	Simpson et al., 1996
Children and adolescent patients	43/212	ORF65 ELISA	Mayama et al., 1998
	33/212	LANA IF	Mayama et al., 1998

[a] BCP-1 and BCBL-1 are HHV-8-infected, EBV-negative PEL cell lines.

contained detectable HHV-8 DNA, and 0 of 11 samples from KS-negative, HIV-negative controls contained detectable HHV-8 DNA (Purvis et al., 1997).

The objectives of the current study were to determine the prevalence of HHV-8 in HIV-negative blood donors, using more sensitive serological tests and PCR, and to determine HHV-8 subtypes in this population by sequencing and analysis of the K1 gene.

3.2 BLOOD DONORS

Blood was collected from 116 blood donors, 98 male and 18 female. A summary of the demographic data is given in Table 3.2. The median age of the donors was 27 years (range, 18-54 years). The majority of the donors were residents of the greater Kampala area, although most of them were born in other areas of Uganda (Table 3.2). Two donors were HIV-positive, and their samples were excluded from further analysis.

3.3 HHV-8 SEROPREVALENCE IN BLOOD DONORS

Two serological tests, ELISA and IF test, were used to analyse plasma from 114 HIV-negative blood donors for antibodies against HHV-8 ORF73 and ORF65 antigens. The raw data are shown in Table 3.3 and are summarized in Table 3.4. The 71 samples that were positive in ORF73 ELISA were also positive in ORF65 ELISA. Of the 51 samples that gave a positive result in the IF test, 43 reacted in the ORF65 ELISA as well and eight in only the IF test. Thus, 97 (85%) samples reacted in at least one of the assays (before confirmation with WB), and 54 (47%) in only one assay. The 46 samples that reacted only in the ELISA (36 with both antigens and 10 with only the ORF65 antigen) were tested by WB to exclude non-specific reactions. Of these, 33 (72%) were confirmed

Table 3.2. Demographic data for blood donors[a]

Donor	Plasma	Cell pellet sample	Specimen date	Gender[b]	Age	Place of residence	Place of birth
001-382	117580-S	99	14-4-98	M	19	Mpigi	Luwero
001-383	117581-U	100	14-4-98	M	27	Makindye	Mbale
001-384	117582-D	101	14-4-98	M	25	Gaba	Zirobwe
001-385	117583-Y	102	14-4-98	M	34	Mulago	Rwashamaire
001-386	117584-W	103	14-4-98	M	26	Kyelima	Entebbe
001-387	117585-U	104	14-4-98	M	30	Kalerwe	Wobulenzi
001-388	117586-C	105	14-4-98	M	38	Makerere	Kakole
001-389	117587-W	106	14-4-98	M	21	Kawempe	Kawempe
001-390	117588-G	107	14-4-98	M	26	Kampala	Bugerere
001-391	117589-O	108	14-4-98	M	24	Kansanga	Mukunyu
001-392	117590-T	109	14-4-98	M	35	Makindye	Mukono
001-394	117746-G	110	15-4-98	M	27	Mulago	Bwera
001-395	117747-D	111	15-4-98	M	25	Wobulenzi	Mawanda Rd
001-396	117748-M	112	15-4-98	M	26	Mulago	Bundibugyo
001-398	117749-I	113	15-4-98	M	31	Kamwokya	Nyenga
001-399	117750-U	114	15-4-98	M	29	Kalangala	Buwali
001-400	117751-N	115	15-4-98	M	20	Nakasero	Gayaza
001-401	117752-R	116	15-4-98	M	30	Ntinda	Kawanda
001-402	117753-C	117	15-4-98	M	21	Ntinda	Ntinda
001-403	117754-V	29	15-4-98	F	41	Namasuba	Nabitumba
001-404	117755-K	30	15-4-98	M	32	Nakulabye	Arua
001-405	117756-J	31	15-4-98	M	30	Zirobwe	Zirobwe
001-406	117757-K	32	15-4-98	M	21	Makerere	Namaliri

001-596	1-596	15	7-5-98	M	26	Maganjo	Katikamu
001-597	1-597	16	7-5-98	M	28	Kibuli	Kabale
001-598	1-598	17	7-5-98	M	26	Nansana	Mukono
001-599	1-599	18	7-5-98	M	32	Kiwonvu	Meru
001-600	1-600	19	7-5-98	F	21	Bukoto	Bukoto
001-601	1-601	20	7-5-98	M	22	Makerere	Makerere
001-602	1-602	21	7-5-98	F	22	Masindi	Masindi
001-603	1-603	22	7-5-98	M	34	Masindi	Mukono
1-604	1-604	23	7-5-98	M	25	Ndeba	Ndeba
1-605	1-605	24	7-5-98	M	48	Masaka	Kaganda
1-607	1-607	25	7-5-98	M	28	Kyengera	Masaka
1-608	1-608	26	7-5-98	M	25	Seeta	Mukono
1-609	1-609	27	7-5-98	M	35	Bwaise	Masaka
1-610	1-610	28	7-5-98	M	24	Masanafu	Masaka
1-409	117849-C	34	16-4-98	M	25	Busega	Masaka
1-410	117850-T	35	16-4-98	M	33	Luzira	Tororo
1-411	117851-H	36	16-4-98	M	27	Kisugu	Tororo
1-412	117852-X	37	16-4-98	M	34	Kitintale	Tororo
1-414	117853-D	38	16-4-98	M	45	Namutamba	Luwonvu
1-415	117854-J	39	16-4-98	M	23	Maganjo	Bombo
1-416	117855-T	40	16-4-98	M	18	Bukoto	Bukoto
1-417	117856-W	41	16-4-98	M	22	Mulago	Rukungiri
1-418	117857-R	42	16-4-98	F	32	Naguru	Kalisizo
1-419	117858-F	43	16-4-98	M	28	Kiboga	Mukono
1-420[a]	117859-U	44	16-4-98	M	30	Mengo	Arua
1-426	118021-I	45	17-4-98	M	32	Kazo	Kamengo
1-427	118022-X	46	17-4-98	M	27	Makerere	Kabale
1-428	118023-N	47	17-4-98	M	29	Kitante	Arua

1-429	118024-G	48	17-4-98	M	54	Makindye	Kabale
1-430	118025-V	49	17-4-98	F	25	Old Kampala	Mbarara
1-431	118026-C	50	17-4-98	M	24	Mpigi	Makerere
1-443	118157-W	51	20-4-98	M	30	Nsambya West	Jinja
1-444	118158-F	52	20-4-98	M	36	Kawempe	Entebbe
1-445	118159-N	53	20-4-98	M	40	Makindye	Mubende
1-446	118160-Y	54	20-4-98	F	22	Kawempe	Kawempe
1-447	118161-R	55	20-4-98	F	29	Kawempe	Bwaise
1-448	118162-G	56	20-4-98	M	41	Kalerwe	Bugerere
1-449	118163-V	57	20-4-98	M	30	Kawempe	Semuto
1-450	118164-O	58[c]	20-4-98	M	28	Kawempe	Kawempe
1-451	118165-V	59	20-4-98	M	25	Kawempe	Kyanawataayi
1-452	118166-I	60	20-4-98	M	25	Kawempe	Wobulenzi
1-453	118167-Q	61	20-4-98	M	20	Kawempe	Mbale
1-454	118168-G	62	20-4-98	M	27	Kawempe	Mbuya
1-455	118169-P	63	20-4-98	M	33	Bugema	Mperwerwe
1-456	118170-T	64	20-4-98	M	41	Kibuye	Kabasanda
1-457	118171-H	65	20-4-98	M	36	Kalerwe	Mulago
1-458	118172-V	66	20-4-98	M	33	Banda	Kayunga
1-459	118173-O	67	20-4-98	M	38	Gayaza	Gayaza
1-460	118174-J	68	20-4-98	F	23	Wampewo	Namirembe
1-461	118175-C	69	20-4-98	M	25	Kasubi	Kasubi
1-462	118176-V	70	20-4-98	M	26	Mukono	Kasubi
1-463	118177-P	71	20-4-98	F	21	Makindye	Rukungiri
1-464	118178-E	72	20-4-98	M	24	Makerere	Ntinda
1-465	118179-A	73	20-4-98	M	39	Bwaise	Butambala
1-469	118244-W	74	21-4-98	M	29	Kansanga	Kansanga
1-472	118245-R	75	21-4-98	M	26	Nakawa	Busia

1-473	118246-M	76	21-4-98	F	20	Naguru	Nagulu
1-474	118357-L	77	22-4-98	M	20	Gaba	Mbarara
1-475	118358-E	78	22-4-98	M	30	Mulago	Iganga
1-476	118359-U	79	22-4-98	M	25	Bunga	Mukono
1-477	118360-T	80	22-4-98	F	25	Kawanda	Kitgum
1-478	118361-D	81	22-4-98	M	32	Bwaise	Mbarara
1-479	118362-W	82	22-4-98	M	32	Namasuba	Namasuba
1-480[a]	118363-M	83	22-4-98	M	29	Kibuye	Namasuba
1-481	118364-G	84	22-4-98	M	19	Mityana	Mityana
1-482	118365-C	85	22-4-98	F	26	Ggaba	Mbale
1-485	118494-Y	86[c]	23-4-98	M	20	Kawempe	Kampala
1-486	118495-W	87	23-4-98	F	24	Kireka	Kampala
1-487	118496-K	88	23-4-98	M	26	Wampewo	Mukono
1-488	118497-G	89	23-4-98	M	20	Kasubi	Kasubi
1-489	118498-M	90	23-4-98	F	26	Mulago	Seeta
1-491	118631-H	91	24-4-98	F	30	Jinja	Makerere
1-492	118632-Y	92	24-4-98	M	19	Mulago	Seeta
1-493	118633-F	93	24-4-98	M	30	Kireka	Nakisunga
1-494	118634-B	94	24-4-98	M	27	Gayaza	Namirembe
1-495	118635-M	95	24-4-98	F	34	Muyenga	Moyo
1-496	118636-G	96	24-4-98	M	29	Kireka	Kasawo
1-497	118637-K	97	24-4-98	M	29	Kampala	Gulu
1-498	118638-S	98	24-4-98	M	32	Luzira	Arua
3-236	126164-Q	118	28-7-98	M	27	Wandegeya	Kagoma
3-237	126165-V	119	28-7-98	M	37	Kasangati	Kasangati
3-238	126166-N	120	28-7-98	M	48	Nansana	Nansana
3-239	126167-T	121	28-7-98	M	23	Kichwamba	Kasese
3-240	126168-Q	122	28-7-98	F	40	Lungujja	Nakatogo

3-241	126169-S	123	28-7-98	M	20	Kasubi	Bweyogerere
3-242	126170-T	124	28-7-98	M	25	Kyengera	Mengo
3-243	126171-V	125	28-7-98	M	36	Mulago Nsoba	Gaba
3-244	126172-N	126	28-7-98	M	23	Kasubi	Namungoona
3-245	126173-W	127	28-7-98	M	25	Kawala	Kawuku
3-246	126174-A	128	28-7-98	M	34	Mulago	Ibanda
3-247	126175-N	129	28-7-98	F	25	Ntinda	Namutamba
3-294	126729-B	132	5-8-98	M	22	Mengo	Busiro
3-295	126730-J	133	5-8-98	M	22	Kalerwe	Kayunga

[a] All blood donors were HIV-1 negative, except two, 1-420 and 1-480, who were HIV-1 positive.

[b] M, male; F, female

[c] Two cell pellet samples that yielded products in PCR analysis.

Table 3.3. HHV-8 serological results[a] for Ugandan blood donors										
					ELISA[b]					
Plasma	Gender[c]	ORF 65	ORF 73	Control[d]	ORF 65	ORF 73	Control	IF[e]	WB	Results
1-596		0.246	0.209	0.154	3+	3+	3+	ns/+ [+]		+
1-597		0.271	0.107	0.047	3+	2+	(+)	ns/+ [+]		+
1-598		0.093	0.147	0.023	3+	3+	-	+		+
1-599		0.051	0.058	0.021	(+) [-]	-	-	-		-
1-600	F	0.214	0.129	0.078	3+	2+	3+	3+		+
1-601		0.064	0.146	0.025	2+	3+	-	-	-	-
1-602	F	0.045	0.033	0.02	(+) [-]	-	-	-		-
1-603		0.198	0.143	0.112	3+	3+	3+	-	+	+
1-604		0.108	0.124	0.024	3+	2+	-	-	-	-
1-605		0.203	0.41	0.049	3+	3+	(+)	-	+	+
1-607		0.052	0.082	0.028	1+	(+)	-	3+		+
1-608		0.386	0.475	0.219	3+	3+	3+	-	+	+
1-609		0.075	0.079	0.028	2+	(+)	-	ns/+ [ns]	+	+
1-610		0.234	0.102	0.043	3+	1+	(+)	ns	+	+
117580-S		0.284	0.187	0.098	3+	3+	3+	ns/- [ns]	+	+
117581-U		0.055	0.088	0.038	1+	1+	-	ns	+	+
117582-D		0.471	0.085	0.028	3+	1+	-	-	+	+
117583-Y		0.035	0.03	0.021	-	-	-	-		-
117584-W		0.063	0.07	0.026	2+	(+)	-	-	+	+
117585-U		0.528	0.038	0.024	3+	-	-	ns/- [ns]	+	+
117586-C		0.068	0.055	0.039	2+	-	-	-	+	+
117587-W		0.083	0.086	0.082	3+	1+	3+	3+		+
117588-G		0.161	0.176	0.046	3+	3+	(+)	-	+	+
117589-O		0.055	0.096	0.042	1+	1+	(+)	3+		+
117590-T		0.028	0.044	0.025	-	-	-	3+		+
117746-G		0.053	0.047	0.029	1+	-	-	-	-	-
117747-D		0.034	0.033	0.023	-	-	-	3+		+
117748-M		0.082	0.072	0.055	3+	(+)	1+	-	+	+
117749-I		0.283	0.199	0.029	3+	3+	-	3+		+
117750-U		0.602	0.342	0.188	3+	3+	3+	3+		+
117751-N		0.116	0.113	0.1	3+	2+	3+	3+		+
117752-R		0.043	0.073	0.035	(+) [-]	(+) [-]	-	ns		-
117753-C		0.143	0.068	0.022	3+	(+)	-	-	-	-
117754-V	F	0.132	0.04	0.021	3+	-	-	3+		+
117755-K		0.103	0.267	0.078	2+	2+	-	-	+	+
117756-J		0.278	0.249	0.162	3+	2+	2+	3+		+
117757-K		0.147	0.257	0.112	3+	2+	(+)	3+		+
117849-C		0.119	0.163	0.052	3+	3+	1+	-	-	-
117850-T		0.028	0.03	0.023	-	-	-	-		-
117851-H		0.033	0.029	0.02	-	-	-	-		-
117852-X		0.336	0.147	0.055	3+	3+	1+	-	+	+
117853-D		0.127	0.089	0.09	3+	1+	3+	2+		+
117854-J		0.074	0.137	0.047	2+	3+	(+)	2+		+
117855-T		0.272	0.056	0.045	3+	-	-	ns	(+)	+
117856-W		0.015	0.015	0.021	-	-	-	ns		-
117857-R	F	0.055	0.086	0.045	(+)	-	-	-	(+)	+
117858-F		0.029	0.135	0.036	-	(+) [-]	-	-		-
118021-I		0.226	0.048	0.035	3+	-	-	3+		+
118022-X		0.204	0.128	0.201	3+	-	3+	3+		+
118023-N		0.295	0.096	0.09	3+	-	(+)	3+		+
118024-G		0.066	0.065	0.144	(+) [-]	-	2+	-		-
118025-V	F	0.063	0.081	0.061	(+) [-]	-	-	-		-
118026-C		0.035	0.04	0.016	-	-	-	-		-
118157-W		0.125	0.079	0.025	3+	(+)	-	-	+	+
118158-F		0.053	0.058	0.019	1+	-	-	-	-	-
118159-N		0.214	0.198	0.075	3+	3+	2+	-	-	-
118160-Y	F	0.193	0.334	0.08	3+	3+	-	ns	+	+
118161-R	F	0.144	0.196	0.064	3+	1+	-	-	+	+
118162-G		0.354	0.478	0.427	3+	3+	3+	2+		+
118163-V		0.106	0.116	0.052	3+	-	-	+		+
118164-O		0.039	0.067	0.039	-	-	-	+		+
118165-V		0.036	0.065	0.038	-	-	-	3+		+
118166-I		0.132	0.117	0.064	3+	2+	2+	2+		+
118167-Q		0.141	0.102	0.5	3+	1+	(+)	+		+
118168-G		0.139	0.092	0.036	3+	1+	-	-	+	+
118169-P		0.041	0.047	0.033	(+) [-]	-	-	-		-
118170-T		0.282	0.103	0.028	3+	1+	-	-	+	+
118171-H		0.115	0.313	0.045	3+	3+	(+)	3+		+
118172-V		0.295	0.365	0.133	3+	3+	3+	3+		+

118173-O		0.217	0.072	0.028	3+	(+)	-	3+		+
118174-J	F	0.234	0.047	0.059	3+	-	1+	3+		+
118175-C		0.55	0.03	0.025	3+	-	-	-	+	+
118176-V		0.148	0.085	0.03	3+	1+	-	-	-	-
118177-P	F	0.09	0.105	0.036	3+	2+	-	2+		+
118178-E		0.381	0.101	0.042	3+	1+	(+)	+		+
118179-A		0.152	0.111	0.05	3+	2+	1+	-	-	-
118244-W		0.075	0.067	0.028	2+	(+)	-	2+		+
118245-R		0.526	0.105	0.058	3+	2+	1+	2+		+
118246-M	F	0.062	0.056	0.022	1+	-	-	3+		+
118357-L		0.35	0.249	0.09	3+	2+	3+	-	+	+
118358-E		0.092	0.047	0.038	3+	-	-	ns/+ [+]		+
118359-U		0.115	0.094	0.022	3+	1+	-	-	+	+
118360-T	F	0.112	0.06	0.03	3+	-	-	-	+	+
118361-D		0.278	0.306	0.03	3+	3+	-	3+		+
118362-W		0.033	0.038	0.017	-	-	-	+		+
118364-G		0.106	0.093	0.024	3+	1+	-	-	-	-
118365-C	F	0.088	0.057	0.026	3+	-	-	-	ns	-
118494-Y		0.252	0.336	0.047	3+	3+	(+)	-	+	+
118495-W	F	0.058	0.066	0.035	1+	(+)	-	2+		+
118496-K		0.132	0.187	0.033	3+	3+	-	+		+
118497-G		0.375	0.132	0.053	3+	2+	1+	3+		+
118498-M	F	0.081	0.123	0.04	3+	2+	(+)	3+		+
118631-H	F	0.041	0.038	0.024	(+)	-	-	+		+
118632-Y		0.437	0.331	0.085	3+	3+	3+	ns	ns	-
118633-F		0.033	0.036	0.018	-	-	-	-		-
118634-R		0.04	0.057	0.02	-	-	-	-		-
118635-M	F	0.058	0.611	0.031	1+	3+	-	3+		+
118636-G		0.216	0.621	0.034	3+	3+	-	3+		+
118637-K		0.044	0.038	0.025	(+) [-]	-	-	-		-
118638-S		0.244	0.174	0.067	3+	3+	2+	ns	+	+
126164-Q		0.12	0.108	0.103	3+	2+	3+	-	(+)	+
126165-V		0.141	0.066	0.036	3+	(+)	-	-	+	+
126166-N		0.675	0.444	0.104	3+	3+	3+	-	+	+
126167-T		0.248	0.235	0.101	3+	3+	3+	3+		+
126168-Q	F	0.033	0.046	0.026	-	-	-	3+		+
126169-S		0.163	0.07	0.062	3+	(+)	2+	-	ns	-
126170-T		0.035	0.041	0.037	-	-	-	3+		+
126171-V		0.09	0.07	0.022	3+	(+)	-	-	+	+
126172-N		0.042	0.064	0.027	(+) [-]	-	-	ns/- [ns]		-
126173-W		0.64	0.188	0.056	3+	3+	2+	(+)		+
126174-A		0.032	0.053	0.033	-	-	-	3+		+
126175-N	F	0.046	0.043	0.045	(+) [-]	-	(+)	-		-
126729-B		0.099	0.074	0.051	3+	(+)	1+	3+		+
126730-J		0.085	0.09	0.035	3+	1+	-	-	+	+
Footnotes overleaf.										

Footnotes for Table 3.3

[a] Results were scored as explained in the Methods section.

[b] Nine (ORF65) and two (ORF73) samples scored as (+), which were negative in the IF test but a WB was not done, were interpreted as being negative ([-]).

[c] Only females are indicated.

[d] Average value for 10 negative control sera.

[e] All 4 samples scored as ns/+ were positive in the ELISA, 3 were counted as being positive ([+]) because a WB was not done while 1 was counted as being non-specific (ns) because a WB was not done. All samples scored as ns/- were counted as being non-specific; interpreted scores are shown in square brackets.

ns, non-specific or indeterminant result.

Table 3.4. Summary of serological results for the different tests

Test	Number tested	Number positive	Number negative	Non-specific, ns	% positive
ORF65 ELISA	114	89[a]	25	0	79
ORF73 ELISA	114	71[a]	43	0	61
IF[b]	114	51	51	12	45
WB	46	33	10	3	72

[a] Excludes positive samples (9 for ORF65 and 2 for ORF73) that did not react in the IF test but where a WB analysis was not performed. These were counted as being negative (see Table 3.3).

[b] Samples that were positive in the ELISA and scored ns/+ in the IF test were interpreted as being non-specific (ns) if a WB analysis was performed or as being positive if a WB was not done; samples scored ns/-, were counted among the non-specific samples (Table 3.3).

positive, 10 were not confirmed and three were indeterminant. As shown in Fig. 3.1, 76 (67%) and 61 (54%) samples had antibodies in the ELISA (confirmed by WB analysis) to the ORF65 and ORF73 recombinant proteins, respectively. All the 51 positive reactions in the IF test were considered to be specific, indicating that 45% of the samples had antibodies to LANA in the IF test (Fig. 3.1). The overall HHV-8 antibody prevalence (including all samples positive in the IF test plus samples confirmed by WB analysis) was 74% (84/114 HIV-negative blood donors) (Fig. 3.1).

3.3.1 Seroprevalence among male and female donors

To assess whether the predominance of males affected the results, the prevalence rates among the females and males were calculated. These were 73% and 77%, respectively, indicating no difference in the prevalence of antibodies among the two groups (Table 3.5A).

3.3.2 Seroprevalence among different age groups

To assess whether there was a correlation between age and prevalence, the prevalence rates by age were calculated. The donors were divided into five groups as follows: 18-23 (n=26); 24-29 (n=46); 30-35 (n=26); 36-41 (n=12); >42 (n=4). The prevalence rates for each group are shown in Table 3.5B. The results show that the prevalence of antibodies is similar in all five age groups.

3.3.3 Conclusion

The serological results indicate that HHV-8 antibodies are highly prevalent in male and female HIV-1-negative blood donors in Kampala, and all age groups show a similar level of antibody prevalence.

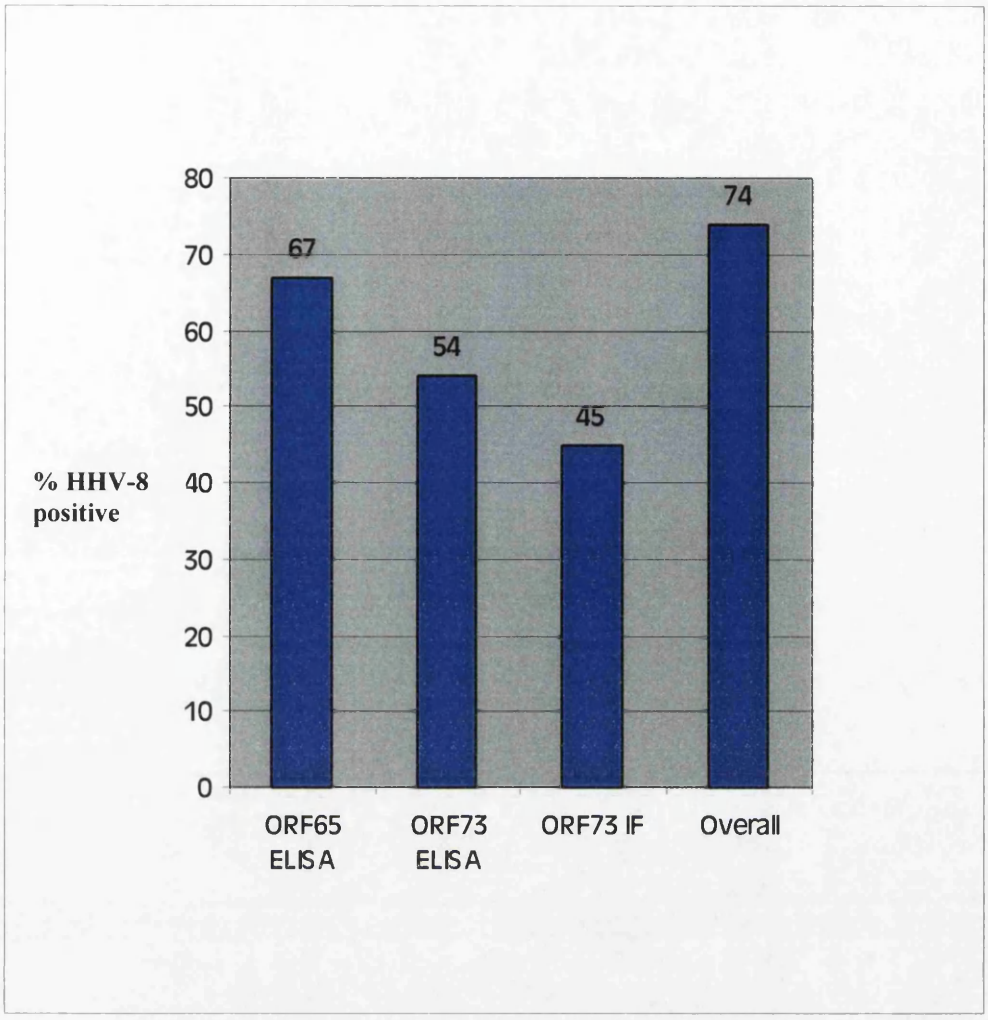


Fig. 3.1. HHV-8 seroprevalence in HIV-1 negative Ugandan blood donors.

The chart shows the seroprevalence rate by test, and the overall seroprevalence, which included all samples positive in the IF test plus samples positive in the WB analysis.

Table 3.5.

A. HHV-8 seroprevalence among male and female blood donors

Gender	Number tested	Number Positive	Number Negative	% Positive
Male	96	70	26	73
Female	18	14	4	77

B. HHV-8 seroprevalence among different age groups

Age group	Number tested	Number Positive	Number Negative	% Positive
18-23	26	18	8	69
24-29	46	34	12	74
30-35	26	20	6	77
36-41	12	9	3	75
>42	4	3	1	75

3.4 FREQUENCY OF HHV-8 DNA IN THE BLOOD DONORS

The objective of this study was to determine the frequency of HHV-8 DNA in PBMC from HIV-negative blood donors by amplification of the ORF26 gene. DNA was extracted from cell pellets consisting of white blood cells. All the samples were initially reacted with ORF26-specific primers (KS1/KS2), and none yielded a product. Then, the samples were amplified by nested PCR with ORF26-specific primers, KS1/KS2 (first round) and NS1/NS2 (second round). The results of these experiments were inconclusive in that the negative controls also had bands of the expected size (data not shown). Finally, the samples were amplified by nested PCR with ORF26-specific primers, KS4/KS5 (first round) and KS1/KS2 (second round). Only 2 (58 and 86) of 114 samples yielded a specific product in these experiments (Fig. 3.2). Both samples came from the same area, Kawempe.

3.5 HHV-8 SUBTYPES IN THE BLOOD DONORS

The aim of this project was to determine the HHV-8 subtypes present in blood donors by K1 PCR and sequence analysis. Initially, attempts were made to amplify the K1 gene of the two HHV-8-positive DNA samples (discussed above) using nested PCR (first round primers, PMC20/PMC21 and second round primers, O1/O2). However, no product was obtained. Attempts to amplify the K1 products of other samples were not made.

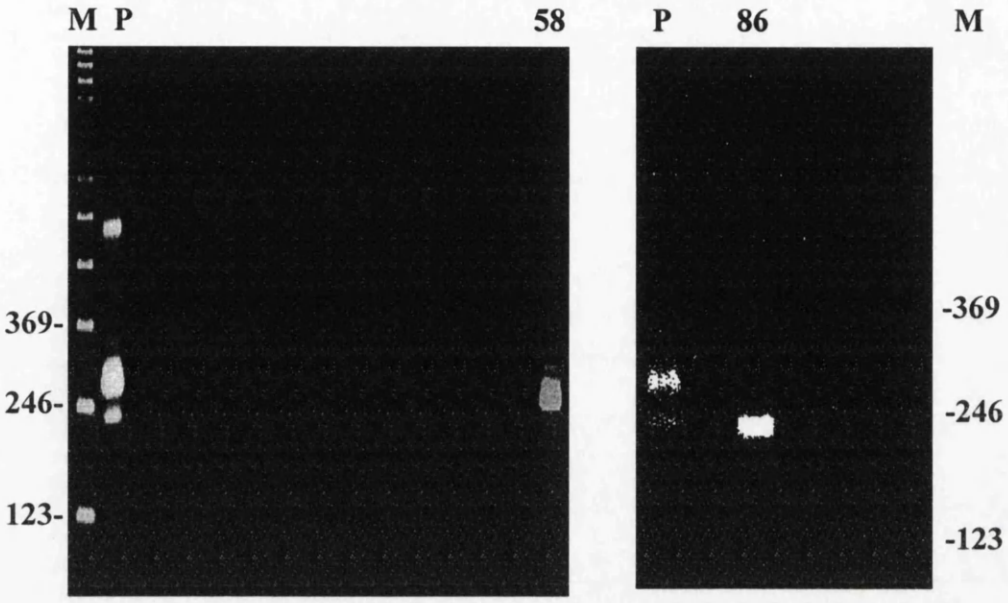


Fig. 3.2. Detection of HHV-8 DNA in blood donors.

EtBr-stained 2% (w/v) agarose gel showing 233 bp products generated by nested PCR from cell pellet DNA samples of blood donors using KS4/KS5 (outer primers) and KS1/KS2 (inner primers). The two samples (58 and 86) that gave products are indicated. P, positive control; M, 123 bp DNA ladder.

3.6 DISCUSSION

3.6.1 HHV-8 seroprevalence in blood donors

ORF65 ELISA detected antibodies in 67% of the blood donors while LANA IF detected antibodies in a lower proportion of individuals (45%). The results of these two tests are not in absolute agreement, as has also been found in other seroprevalence studies involving African samples (Table 3.1; Lennette et al., 1996; Mayama et al., 1998). The two tests seem to complement each other; ORF65 ELISA detected HHV-8 antibodies in 33 individuals that did not react in the LANA IF test, while LANA IF test detected antibodies in eight individuals that did not react in the ORF65 ELISA. Combined, the two tests detected antibodies in a high percentage of blood donors (74%), indicating that HHV-8 infection is widespread in Kampala, and probably in Uganda as a whole. This result falls within the range of seroprevalence rates reported by other studies for the general Ugandan population (Table 3.1). However, it was higher than the rate reported for blood donors (Table 3.1; Chatlynne and Ablashi, 1999). This could be explained by the fact that the combined tests used in the current study are more sensitive than the single test (whole virus ELISA) that was used in the previous study. Mayama et al. (1998) reported that HHV-8 transmission occurs during early childhood in Uganda and reaches adult levels (in approx. 50% of the children) before the age of puberty. The high prevalence of antibodies in the adult blood donor population studied here may be a reflection of a lifetime presence of HHV-8 infection in the majority of the Ugandan population.

Although the blood donor population in this study was predominantly male, the results show that HHV-8 antibodies are distributed in virtually equal proportions among men and women. A study of HHV-8 distribution in blood donors in Italy reported similar findings (Calabrò et al., 1998), whereas other studies have

reported that men are twice as likely to be seropositive than women (Luppi et al., 1998; Manns et al., 1998). This disparity may be explained by differences in HHV-8 seroprevalence in the different areas, whereby HHV-8 antibodies tend to be equally distributed among men and women in areas of high seroprevalence.

3.6.2 Frequency of HHV-8 DNA in blood donors

Previous studies have reported that HHV-8 is not, or only infrequently, detected by PCR in the peripheral blood of healthy donors (Ambroziak et al., 1995; Bigoni et al., 1996; Brambilla et al., 1996; Humphrey et al., 1996; Lefrere et al., 1996; Moore et al., 1996; Whitby et al., 1995). Consistent with these results, HHV-8 was detected in only 1.7% of cell pellet DNA samples of the blood donors in this study. However, this finding was surprising in view of the high seroprevalence detected in this population. HHV-8 DNA has been detected in 10% of PBMCs from HIV-negative individuals in The Gambia, an area of high HHV-8 seroprevalence (89%) like Uganda. Thus it is possible that, with further optimization of the experimental conditions, the level of detection could be increased.

CHAPTER 4
RESULTS AND DISCUSSION-II
CHARACTERIZATION OF THE K1 GENE

4.1 INTRODUCTION

HHV-8 strains from various parts of the world fall into five groups (A/I, B/IV, C/II, D/III, E) (Biggar et al., 2000; Cook et al., 1999; Kasolo et al., 1998; Lacoste et al., 2000a,b; Meng et al., 1999, 2001; Zong et al., 1999). The majority of HHV-8 subtypes from the USA and Europe (A and C) are relatively closely related, whereas the B, D and E strains identified from Africa, Pacific Islands and Brazilian Amerindians, respectively, are phylogenetically more distant from these subtypes. Thus far, 13 HHV-8 strains from Uganda have been published (Table 4.1); two (Ugd1 and Ugd2) are part of this study. Apart from Ugd2, all other samples (with the exception of UgR1 whose source is not clear) came from HIV-positive individuals. The strains belong to three main HHV-8 subtypes, A (2/13), A5 (3/13), B (7/13) and C (1/13), subtype B being predominant. Kasolo et al. (1998) found that 15 of 15 HHV-8-positive children in Zambia (with or without KS) had subtype A5 K1 genomes in their peripheral lymphocytes. All five strains from Zambian AIDS patients that Zong et al. (1999) characterised were subtype B. Recently, Lacoste et al. (2000a) published a large study including 32 strains of African origin; 21 from West and Central Africa. The majority of these strains were either subtype B (13 cases) or variant A5 (11 cases), the rest were subtype A or C. One strain from Central African Republic (K1-8/Dem) had a K1 sequence with characteristics of both subtypes A and C, and another one (K1-43/Ber) belonged to the A/C lineage but also had characteristics of neither the C nor the A subtype. All these studies noted no obvious correlation between K1 subtypes and the different forms of KS (clinical and epidemiological).

Table 4.1. Published HHV-8 strains from Uganda

Strain	Gender	Source	K1 subtype	Reference
ST1	F	AIDS KS biopsy	B	Zong et al., 1999
ST2	F	AIDS KS biopsy	B	Zong et al., 1999
Ugd1	M	AIDS KS biopsy	B	Cook et al., 1999
Ugd2	M	Endemic KS biopsy	B	Cook et al., 1999
Ug52	M	HIV positive, no KS PBMC	B	Cook et al., 1999
Ug81	M	AIDS KS PBMC	B	Cook et al., 1999
Ug1	F	AIDS KS PBMC	B	Meng et al., 1999
Ug111	M	AIDS KS PBMC	A	Cook et al., 1999
Ug4	M	AIDS KS PBMC	A	Meng et al., 1999
Ug374	M	HIV positive, no KS, PBMC	A5	Cook et al., 1999
Ug2	M	HIV positive PBMC	A5	Meng et al., 1999
Ug3	M	HIV positive PBMC	A5	Meng et al., 1999
UgR1	unknown	PBMC	C	Cook et al., 1999

Virtually all the Ugandan strains whose K1 gene has been determined came from HIV-positive individuals. The objectives of this study were to characterise the K1 gene in samples from both HIV-positive and HIV-negative KS patients from Uganda and to evaluate the correlation between K1 subtypes and the type of KS, ethnicity and place of residence of the patients.

4.2 KS PATIENTS

Thirty KS patients (25 male, 5 female) accessed at UCI were the source of the HHV-8-positive DNA samples used in this study and in the studies reported in chapters five and six below. A summary of the demographic and clinical data is shown in Table 4.2. The median age was 38 years (range, 16-70 years). All patients came from the southern part of the country and represent 12 tribes. Almost half (14) of the patients belong to the Ganda tribe, which is the predominant tribe in the Kampala area. Twenty patients were HIV-1 positive and 10 were negative. The majority had nodular KS lesions involving the lower, upper or both extremities. Several of the HIV-positive patients had lesions covering the whole body surface, and some also had generalised KS with lymphadenopathy. Patient Ugd12 had clinically aggressive non-AIDS-associated (endemic) KS. This patient was tested several times for HIV antibodies by Western blot analysis but was found to be negative in all instances.

Plasma from 19 patients tested for HHV-8 antibodies (as described in the previous section) were all positive with both the ELISA and the IF test. The 11 plasma samples that were not tested had been inadvertently left in Uganda.

Table 4.2. Demographic and clinical data for KS patients

Sample[a]	K1 subtype	Age (yr)	Gender[b]	Tribe	Residence (city or district)	HIV status[c]	Clinical presentation	
							Type of lesion	Location of lesion
Ugd1	B	65	M	Ganda	Mukono	+	patches/plaques	upper extremities
Ugd2	B	53	M	Nyori	Iganga	-	nodules	lower/ upper extremities
Ugd3	B	31	M	Nyarwanda	Kampala/then Rwanda	+	nodules	right foot
Ugd4	A5	42	M	Nyankole	Ntungamo	-	nodules	lower/ upper extremities
Ugd7	B	47	M	Nyarwanda	Mubende	+	nodules/plaques	lower extremities/ oral cavity
Ugd10	B	70	M	Mtoro	Kabarole	-	nodules	lower/ upper extremities
Ugd12	A5	54	F	Ganda	Mpigi	-[d]	aggressive/plaques/ lymphadenopathy	whole body
Ugd13	B	29	M	Ganda	Kampala	+	plaques	lower extremities
Ugd15	B	38	M	Lugbara	Mbarara	+	nodules/plaques	extremities/trunk
Ugd16	A5	35	M	Ganda	Mpigi	+	nodules	lower extremities
Ugd18	A5	65	M	Ganda	Masaka	-	nodules	lower/ upper extremities
Ugd19	B	16	M	Jopadhola	Tororo	-	nodules	lower/ upper extremities
Ugd21	B	30	M	Samia	Mukono	+	nodules/plaques	whole body
Ugd23	C	43	M	Mfumbira	Jinja	+	nodules	lower extremities
Ugd24	A5	52	M	Konjo	Kasese	-	nodules	lower extremities
Ugd26	B	61	M	Ganda	Mukono	-	nodules	lower extremities
Ugd29	B	67	M	Ganda	Mpigi	-	nodules	nodules
Ugd5		38	F	Ganda	Mpigi	+	plaques	whole body/ oral cavity
Ugd6		32	F	Luo	Kampala	+	nodules	trunk/forearm/ hard palate
Ugd8		33	M	Ganda	Mpigi	+	nodules	lower extremities/ oral cavity
Ugd9		29	M	Nyoro	Kampala	+	nodules/plaques	whole body/ oral cavity
Ugd11		28	M	Nyankole	Bushenyi	+	nodules	head/extremities
Ugd14		66	M	Ganda	Kampala	-	nodules	lower/ upper extremities
Ugd17		45	M	Ganda	Mpigi	+	plaques/ lymphadenopathy	generalised
Ugd20		29	M	Ganda	Kampala	+	nodules/plaques	whole body
Ugd22		43	M	Ganda	Mpigi	+	plaques/patches	lower extremities
Ugd25		30	F	Jopadhola	Kampala	+	plaques	chest
Ugd27		37	M	Ganda	Mpigi	+	nodules/plaques/ lymphadenopathy	generalised KS
Ugd28		31	M	Langi	Luwero	+	nodules/plaques	generalised KS
Ugd30		20	F	Nyarwanda	Mubende	+	records not available	records not available

[a] Samples whose K1 subtypes are known are listed first.

[b] M, male; F, female.

[c] +, HIV-1 positive; -, HIV-1 negative.

[d] Clinically aggressive, non-AIDS-associated KS.

4.3 IDENTIFICATION OF HHV-8-POSITIVE DNA SAMPLES

To identify HHV-8-positive samples, all tumour DNA samples were amplified by PCR with HHV-8 ORF26-specific primers, initially KS1/KS2 primers and later KS4/KS5 primers. All samples yielded a product of the expected sizes with both primer sets, except Ugd14, which did not yield a product with KS1/KS2 (Fig. 4.1A). This sample yielded a relatively faint product with KS4/KS5 (Fig. 4.1B). Ugd5 also yielded a very faint band on amplification with KS1/KS2. As a control for DNA extraction, the β-globin gene was amplified for samples Ugd1-Ugd16 using primers PCO_3 and P64 (Fig. 4.2). This also allowed rough estimation of DNA concentration in the samples. Ugd5 yielded a faint band compared to other samples, indicating that the DNA concentration in this sample was relatively low. This could explain the faint band observed on amplification with KS1/KS2 (Fig. 4.1A). Ugd14, on the other hand, had a β-globin-specific band of similar intensity to that of other samples (with the exception of Ugd5), indicating similar levels of DNA concentration. Ugd17-Ugd30 were not amplified for β-globin because it was judged that they contain similar concentrations of DNA as the tested samples (with the exception of Ugd5), inasmuch as they yielded signals of similar intensity in the ORF26 PCR analysis.

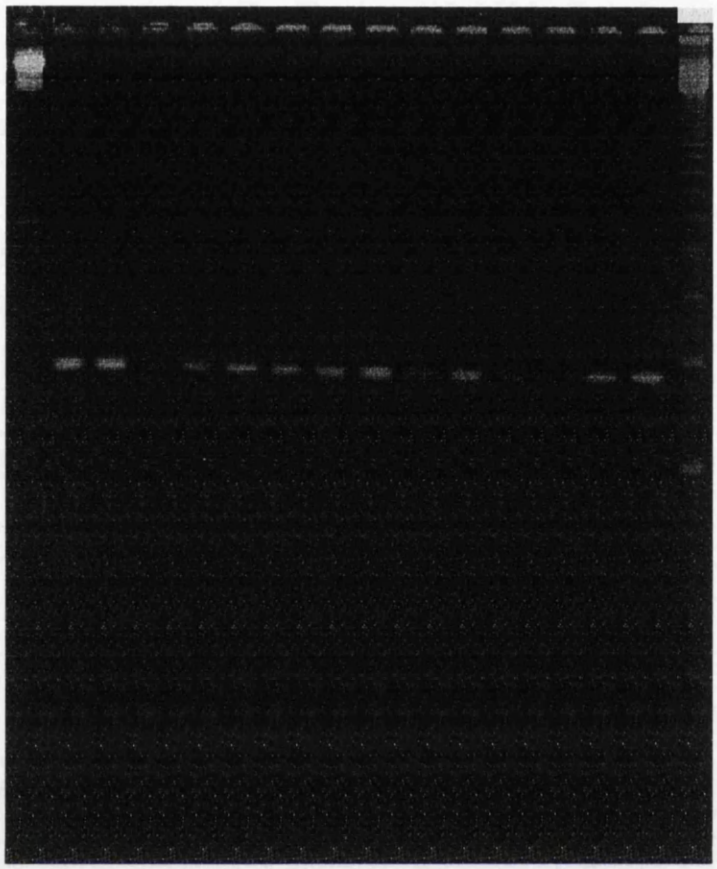
Cell pellet DNA samples from 15 KS patients (Ugd3-Ugd17) were subjected to PCR with ORF26-specific primers (KS1/KS2). None yielded a product. However, three of these samples (Ugd7, Ugd11 and Ugd16) yielded a product in nested PCR with KS4/KS5 as outer primers and KS1/KS2 as inner primers (Fig. 4.3).

The level of detection of HHV-8 DNA by PCR in PBMC of KS patients has been reported to be 50-60% (Whitby et al., 1995). Another study involving Ugandan KS patients reported a higher level of detection (84%) in PBMC using

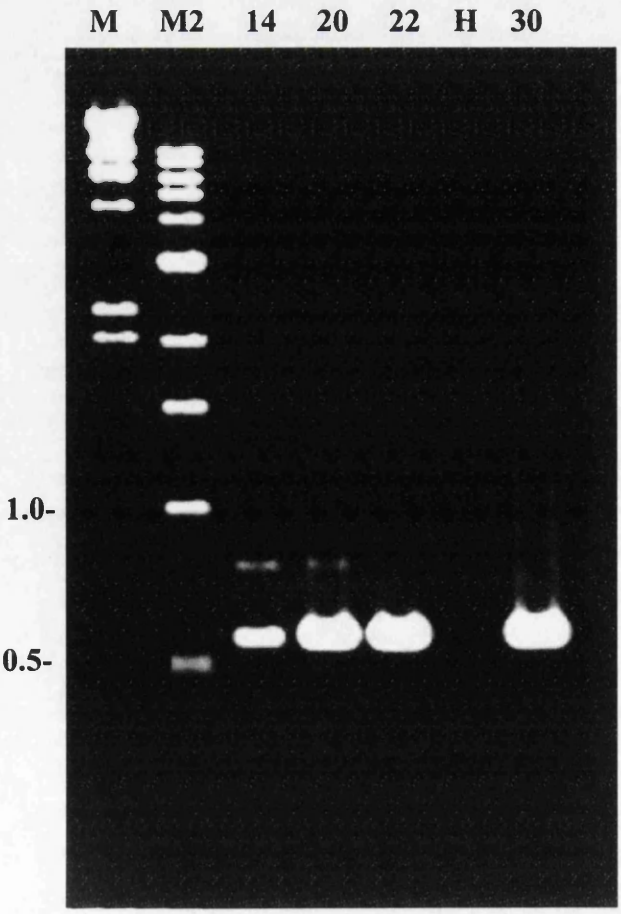
Fig. 4.1. Detection of HHV-8 DNA in KS skin tumour DNA samples.

EtBr-stained 2% (A) and 1% (B) (w/v) agarose gels showing, respectively, 233 and 571 bp products of samples amplified with ORF26-specific primers, KS1/KS2 (A) and KS4/KS5 (B). Representative PCR products are shown. The numbers at the top of the lanes denote sample names without the Ugd-prefix, e.g. 3 stands for Ugd3. H, negative (water) control; M, lambda DNA-*Hin*d III digest markers; M1 and M2 are 123 bp and 1 kbp DNA ladders, respectively. The markers to the right of panel A and to the left of panel B are in bp and kbp, respectively.

A



B



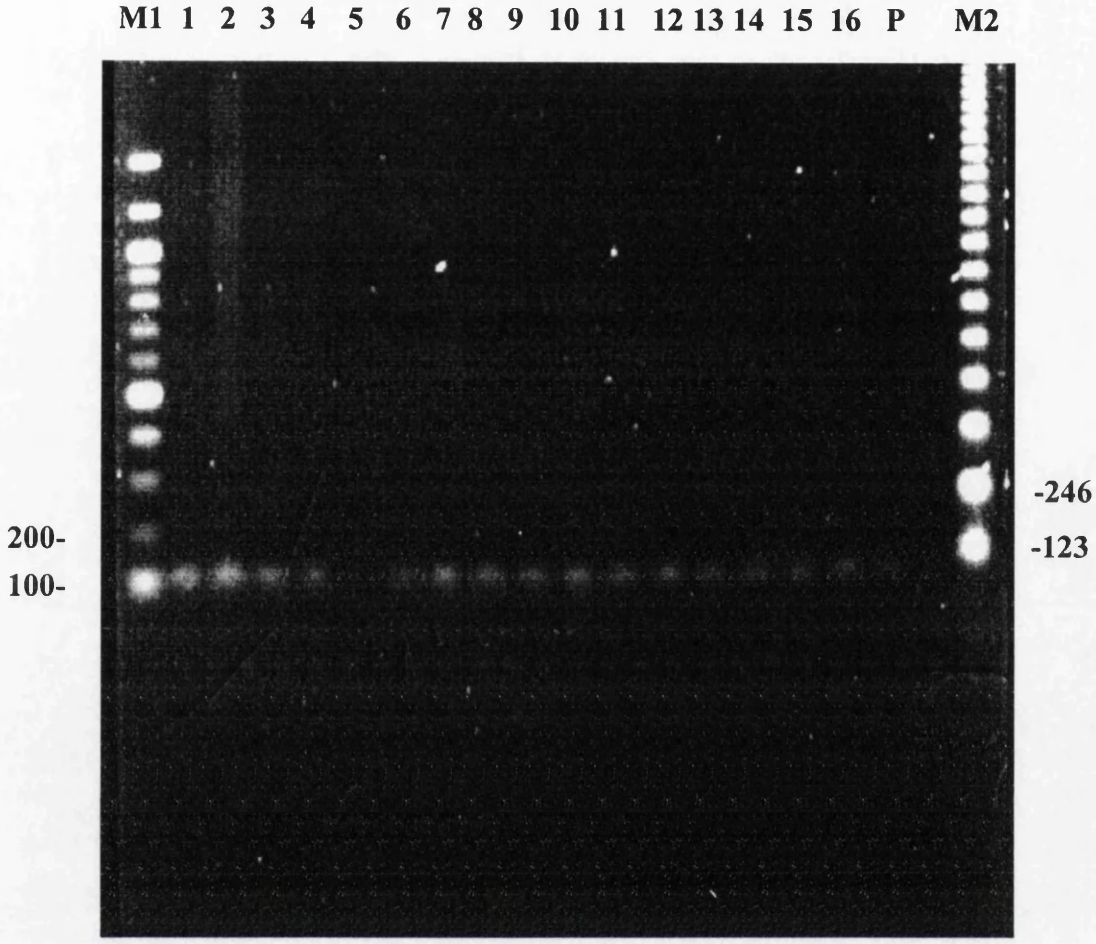


Fig. 4.2. β-globin gene products.

EtBr-stained 1% (w/v) agarose gel showing 110 bp products generated from KS skin tumour DNA samples using PCO_3/P64. The numbers at the top of the lanes denote sample names without the Ugd-prefix, e.g. 3 stands for Ugd3. P, positive control; M1 and M2 are 100 bp and 123 bp DNA ladders, respectively.

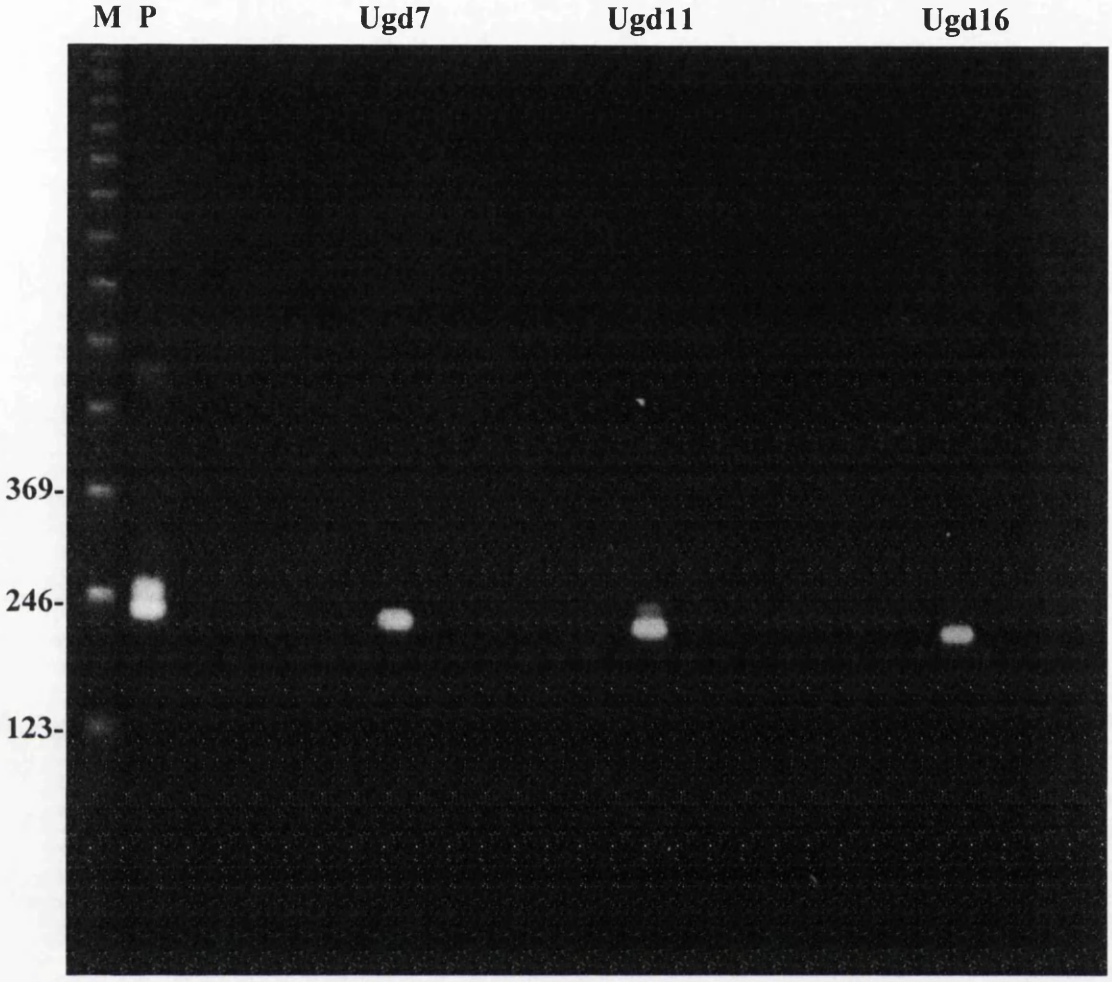


Fig. 4.3. Detection of HHV-8 DNA in KS cell pellet DNA samples.

EtBr-stained 1.5% (w/v) agarose gel showing 233 bp products generated by nested PCR with KS4/KS5 (outer primers) and KS1/KS2 (inner primers). The samples that gave products are indicated at the top of the respective lanes. P, positive control; M, 123 bp DNA ladder.

a combination of PCR and Southern hybridization (Purvis et al., 1997). Therefore, the low frequency of HHV-8 DNA detected in the white blood cells of the patients in this probably indicates suboptimal experimental conditions rather than absolute prevalence.

4.4 PCR AMPLIFICATION OF THE K1 GENE

The entire K1 gene was amplified as a 1277 bp or 874 bp product. PCR and cloning were successful for 17 (of 30) samples; representative PCR products are shown in Fig. 4.4. Ugd1 and Ugd2 (Fig. 4.4A) gave stronger signals compared to samples that were collected subsequently (Fig. 4.4B). Since both sets of samples had similar DNA concentrations as revealed by β-globin analysis (Fig. 4.2), the discrepancy in signals is most likely due to presence of higher HHV-8 copy numbers in Ugd1/2 than in other samples. The PCR products of Ugd4/12/24 were obtained by nested PCR (874 bp; Fig. 4.4C). The band of Ugd12 migrated slightly faster than bands of expected size. The PCR products were cloned before sequencing. Insufficient amounts of purified PCR products were obtained for some of the samples shown in Fig. 4.4B, so they were not cloned.

4.5 PHYLOGENETIC ANALYSIS OF THE K1 GENE

The 17 DNA (and amino acid) sequences obtained in this study together with 35 sequences from previous studies (Cook et al., 1999; Lacoste et al., 2000a; Zong et al., 1999) were aligned as shown in Fig. 4.5A. The translated amino acid sequences were also aligned (Fig. 4.5B). TKS10 (Table 2.1) has a 39-bp (13-aa) duplication of sequences immediately upstream of the insertion (Fig. 4.5A,B). Phylogenetic analysis of the DNA sequences is presented in Fig. 4.6.

Fig. 4.4. PCR amplification of the entire K1 gene.

EtBr-stained 1% (w/v) agarose gels showing representative PCR products amplified from KS skin tumour DNA samples with K1-specific primers, (A, B; 1277bp) PMC20/21 and (C; 874 bp), nested PCR with PMC20/21 (outer primers) and O1/O2 (inner primers). Numbers at the top of the lanes in (B) and (C) represent sample names without the Ugd-prefix; P, positive control (Ugd2). Samples in (B) were done in duplicates. M1 and M2 are lambda DNA-*Hin*d III digest markers and 1 kbp DNA ladder, respectively.

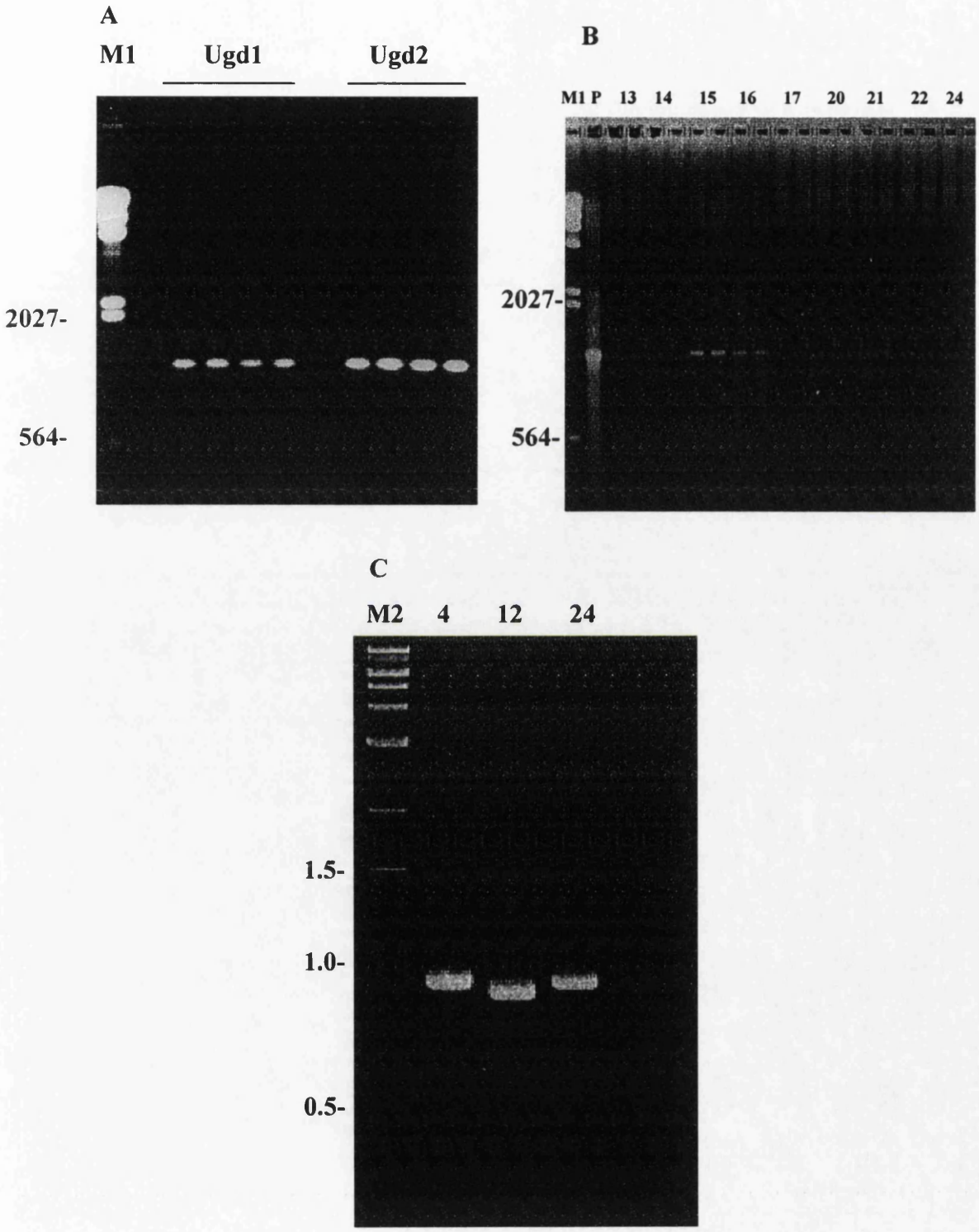

A
M1
Ugd1
Ugd2
2027-
564-
B
M1 P 13 14 15 16 17 20 21 22 24
2027-
564-
C
M2 4 12 24
1.5-
1.0-
0.5-

Fig. 4.5. Alignments of K1 DNA and amino acid sequences of 52 HHV-8 strains.

The alignments include almost complete K1 DNA (A) or amino acid (B) sequences, 17 from this study and 35 from previous studies (Cook et al., 1999; Zong et al., 1999; Lacoste et al., 2000a). Prefixes on sequence names denote K1 subtypes. The alignments run in the orientation of the genomic sequence; nucleotides 1 and 879 are equivalent to 120 and 959, respectively, in the BC-1 genome. Fifteen base pairs were omitted at each end of the DNA sequences, while 5 and 4 amino acids were omitted at the N-terminus and C-terminus, respectively, of the protein sequences.

A

```
                1                                                          60
   d1-tks10  GTCTGCAGTC TGGCGGTTTG CTTTCCAGGA CTATTGAGCC TTCATCTGTC AGTGTCTCCA
    d2-zks3  GTCTGCAATC TGGCGGTTTG CTTTCCAGGA CTATTGAGCC TTCATCTGCC AGCGTTTCCA
     b-ugd1  CTCTGCTGTT TGGTGGTTTG GTTTCCAAAA CTATTGAGCC TTCATCTGCC ATCGTTTCCA
    b-ugd26  CTCTGCAGTT TGGTGGTTTG GTTTCCAAAA CTATTGAGCC TTCATCTGCC ATCGTTTCCA
    b-ugd19  CTCTGCAGTT TGGTGGTTTG GTTTCCAAAA CTATTGAGCC TTCATCTGCC ATCGTTTCCA
     b-ugd7  CTCTGCAGTT TGGTGGTTTG CTTTCCAAAA CTATTGAGCC TTCATCTGCC ATCGTTTCCA
     b-ugd2  CTCTGCAGTT TGGTGGTTTG GTTTCCAAAA CTATTGAGCC TTCATCTGCC ATCGTTTCCA
    b-ugd21  CTCTGCAGTT TGGTGGTTTG GTTTCCACAA CTATTGAGCC TTCATCTGCC ATCGTTTCCA
    b-ugd10  GTCTGCAGTT TGCTGGTTTG CTTTCCAAAA CTATTGAGCC TTCATCTGCC ATCGTTTCCA
   b-ukma24  CTCTACGGTT TTCTGGTTTG CTTTCCAAAA CTATTGAGCC TTCATCTGCC ATCGTTTCCA
    b-ugd29  GTCTCCAGTT TGCTGGTTTG CTTTCCAAAA CTATTGAGCC TTCATCTGCC ATCGTTTCCA
     b-ug52  GTCTGCAGTC TGGCGGTTTG CTTTCCAAAA CTATTGAGCC TTCATCTGCC ATCGTTTCCA
     b-g413  GTCTGCAGTC TGGCGGTTTG CTTTCCAAAA CTATTGAGCC TTCATCTGCC ATCGTTTCCA
     b-g482  GTCTGCAGTC TGGCGGTTTG CTTTCCAAAA CTATTGAGCC TTCATCTGCC ATCGTTTCCA
    b-ugd13  GTCTGCAGTT TGCTGGTTTG CTTTCCAAAA CTATTGAGCC TTCATCTGCC ATCGTTTCCA
     b-ugd3  GTCTGCAGTT TGCTGGTTTG CTTTCCAAAA CTATTGAGCC TTCATCTGCC ATCGTTTCCA
    b-ugd15  GTCTGCAGTT TGCTGGTTTG CTTTCCAAAA CTATTGAGCC TTCATCTGCC ATCGTTTCCA
     b-g71   GTCTGCAGTT TGCTGGTTTG CTTTCCACAA CTATTGAGCT TTAATCTGCC ATCGTTTCCA
     b-g91   CTCTGCAATT TGGTGGTTTG CTTTCCAAAA CTATTGAGCC TTCATCTGCC ATCGTTTCCA
     c-gk17  GTCTGCAGTC TGGCGGTTTG CTTTCGAGGA CTATTAAGCC TTTATCTGCA ATCGTCTCCA
     c-iap3  GTCTGCAGTC TGGCGGTTTG CTTTCGAGGA CTATTAAGCC TTTATCTGCA ATCGTCTCCA
     c-erla  GTTTGCAGTC TGGCGGTTTG CTTTCGAGGA CTATTAAGCC TTTCTCTGCA ATCGTCTCCA
    c-icam1  GTTTGCAGTC TGGCGGTTTG CTTTCGAGGA CTATTAAGCC TTTCTCTGCA ATCGTCTCCA
    c-ukma1  GTTTGCAGTC TGGCGGTTTG CTTTCGAGGA CTATTAAGCC TTTCTCTGCA ATCGTCTCCA
     c-iap2  GTTTGCAGTC TGGCGGTTTG CTTTCGAGGA CTATTAAGCC TTTCTCTGCA ATCGTCTCCA
     c-bbg1  GTCTGCAGTC TGGCGGTTTG CTTTCGAGGA CTATTAAGCC TTTCTCTGCA ATCGTCTCCA
    c-ukma8  GTTTGCAGTC TGGCGGTTTG CTTTCGAGGA CTATTAAGCC TTTCTCTGCA ATCGTCTCCA
      c-bc2  GTTTGCAGTC TGGCGGTTTG CTTTCGAGGA CTATTAAGCC TTTCTCTGCA ATCGTCTCCA
    c-ukma3  GTTTGCAGTC TGGCGGTTTG CTTTCGAGGA CTATTAAGCC TTTCTCTGCA ATCGTCTCCA
     c-ive1  GTAGGCAGTC TGGCGGTTTG CTTTCAAGGA CTATTAAGCC TTTCTCTGCA ATCGTCTCCA
    c-ukb22  GTTTGCAGTC TGGCGGTTTG CTTTCGAGGA CTATTAAGCC TTTCTCTGCA ATCGTCTCCA
     c-gk18  GTTTGCAGTC TGGCGGTTTG CTTTCGAGGA CTATTAAGCC TTTCTCTGCA ATCGTCTCCA
  c-k1-8-dem GTCTGCAGTC TGGCGGTTTG CTTTCGAGGA CTATTAAGCC TTTATGTGCA ATCGTCTCCA
    c-ugd23  GTCTGCAGTC TGGCGGTTTG CTTTCGAGGA CTATTAAGCC TTTATGTGCA ATCGTCTCCA
    a-ugd16  GTCTGCAGTC TGGCGGTTTG CTTTCAAGGA CTATTAAGCC TTTATCTGCA ATCGTCTCCA
    a-ugd18  GTCTGCAGTC TGGCGGTTTG CTTTCAAGGA CTATTAAGCC TTTATCTGCA ATCGTCTCCA
    a-ugd24  GTCTGCAGTC TGGCGGTTTG CTTTCAAGGA CTATTAAGCC TTTATCTGCA ATCGTCTCCA
    a-ug374  GTCTGCAGTC TGGCGGTTTG CTTTCGAGGA CTATTAAGCC TTTATCTGCA ATCGTCTCCA
     a-ugd4  GTCTGCAGTC TGGCGGTTTG CTTTCGAGGA CTATTAAGCC TTTATCTGCA ATCGTCTCCA
    a-bcbl1  GTCTGCAGTC TGGCGGTTTG CTTTCGAGGA CTATTAAGCC TTTATGTGCT ATCGTCTCCA
    a-uka13  GTCTGCAGTC TGGCGGTTTG CTTTCGAGGA CTATTAAGCC TTTCTCTGCT ATCGTCTCCA
    a-uka21  GTCTGCAGTC TGGCGGTTTG CTTTCGAGGA CTATTAAGCC TTTCTCTGCT ATCGTCTCCA
    a-ukma4  GTCTGCAGTC TGGCGGTTTG CTTTCGAGGA CTATTAAGCC TTTCTCTGCT ATCGTCTCCA
      a-bc1  GTCTGCAGTC TGGCGGTTTG CTTTCGAGGA CTATTAAGCC TTTCTCTGCT ATCGTCTCCA
     a-ema7  GTCTGCAGTC TGGCGGTTTG CTTTCGAGGA CTATTAAGCC TTTCTCTGCT ATCGTCTCCA
    a-bcblr  GTCTGCAGTC TGGCGGTTTG CTTTCGAGGA CTATTAAGCC TTTCTCTGCT ATCGTCTCCA
    a-ukb12  GTCTGCAGTC TGGCGGTTTG CTTTCGAGGA CTATTAAGCC TTTCTCTGCT ATCGTCTCCA
    a-ug111  GTCTGCAGTC TGGCGGTTTG CTTTCGAGGA CTATTAAGCC TTTCTCTGCT ATCGTCTCCA
     a-iap1  GTCTGCAGTC TGGCGGTTTG CTTTCGAGGA CTATTAAGCC TTTCTCTGCT ATCGTCTCCA
    a-ukc12  GTCTGCAGTC TGGCGGTTTG CTTTCGAGGA CTATTAAGCC TTTCTCTGCT ATCGTCTCCA
c-k1-43-ber GTCTGCAGTC TGGCGGTTTG CTTTCCAGGA CTATTAAGCC TTTATGTGCA ATCGTCTCCA
    a-ugd12  GTCTGCAGTC TGGCGGTTTG CTTTCGAGGA CTATTAAGCC TTTATCTGCA ATCGTCTCCA
  Consensus  -T---C--T- T---GGTTTG -TTTC-A--A CTATT-AGC- TT--T-TG-- A--GT-TCCA
```

```
            61                                                          120
   d1-tks10 CAATTCTGCC CTGCAGTGCT TTCTACGTCT TACACGTTGA CCTGTCTCTC TGATGCATCC
    d2-zks3 CCATTGTGCC CTGGAGTGCT TTCTACGAAT TACACGTTGA CCTGTCTCTC TGATGCATCC
     b-ugd1 CATTTGTGCC CTGGAGTGCT TTTCACGCCT TACACGTTGA CTTGTCCGTC TAACAGATCC
    b-ugd26 CATTTGTGCC CTGGAGTGCT TTTCACGCCT TACACGTTGA CTTGTCCGTC TAACAGATCC
    b-ugd19 CATTTGTGCC CTGGAGTGCT TTTCACGCCT TACACGTTGA CTTGTCCGTC TAACAGATCC
     b-ugd7 CATTTGTGCC CTGGAGTGCT TTTCACGCCT TACACGTTGA CTTGTCCGTC TAACAGATCC
     b-ugd2 CATTTGTGCC CTGGAGTGCA TTTCACGCCT TACACGTTGA CTTGTCCGTC TAACAGATCC
    b-ugd21 CATTTGTGCC CTGGAGTGCA TTTCACGTCT TACACGTTGA CTTGTCCGTC TAACAGATCC
    b-ugd10 CATTTGTGCC CTGGAGTGAT TTCCACGCCT TACACGTTGA CCTGTCCGTC TAACAGATCC
   b-ukma24 CATTTGTGCC CTGGAGTGAT TTCCACGCCT TACACGTTGA CCTGTCCGTC TAACAGATCC
    b-ugd29 CCTTTGTGCC CTGGAGTGAT TTCCACGCCT TACACGTTGA CCTGTCCGTC TAACAGATTC
     b-ug52 CATTTGTGCC CTGGAGTGAT TTCCACGCCT TACACGTTGA CCTGTCCGTC TACCAGATCC
     b-g413 CATTTGTGCC CTGGAGTGAT TTCCACGCCT TACACGTTGA CCTGTCCGTC TAACAGATCC
     b-g482 CATTTGTGCC CTGGAGTGCT TTCCACGCCT TACACGTTGA CCTGTCCGTC TAACAGATCC
    b-ugd13 CATTTGTGCC CTGGAGTGGT TTCCACGCCT TACACGTTGA CCTGTCCGTC TAACAGATCC
     b-ugd3 CATTTGTGCC CTGGAGTGGT TTCCACGCCT TACACGTTGA CCTGTCCGTC TAACAGATCC
    b-ugd15 CATTTGTGCC CTGGAGTGGT TTCCACGCCT TACACGTTGA CCTGTCCGTC TAACAGATCC
      b-g71 CATTTGTGCC CTGGAGTGAT TTCCACGCCT TACACGTTGA CCTGTCCGTC TAACAGATCC
      b-g91 CATTTGTGCC CTGGAGTGAT TTCCACGCCT TACACGTTGA CCTGTCCGTC TAACAGATCC
     c-gk17 AATATGTGCC CTGGAGTGAT TTCAACGCCT TACACGTTGA CCTGTCCGTC TGGTACATCC
     c-iap3 AATATGTGCC CTGGAGTGAT TTCAACGCCT TACACGTTGA CCTGTCCGTC TAATACATCC
     c-erla AATTTGTGCC CTGGAGTGAT TTCAACGCCT TACACGTTGA CCTGTCCGTC TAATACATCC
    c-icam1 AATTTGTGCC CTGGAGTGAT TTCAACGCCT TACACGTTGA CCTGTCCGTC TAATACATCC
    c-ukma1 AATTTGTGCC CTGGAGTGAT TTCAACGCCT TACACGTTGA CCTGTCCGTC TAATACATCC
     c-iap2 AATTTGTGCC CTGGAGTGAT TTCAACGACT TACACGTTGA CCTGTCCGTC TAATACATCC
     c-bbg1 AATTTGTGCC CTGGAGTGAT TTCAACGCCT TACACGTTGA CCTGTCCGTC TAATACATCC
    c-ukma8 AATTTGTGCC CTGGAGTGAT TTCAACGCCT TACACGTTGA CCTGTCCGTC TAATACATCC
      c-bc2 AATTTGTGCC CTGGAGTGAT TTCAACGCCT TACACGTTGA CCTGTCCGTC TAATACATCC
    c-ukma3 AATTTGTGCC CTGGAGTGAT TTCAACGCCT TACACGTTGA CCTGTCCGTC TAATACATCC
     c-ive1 AATTTATGCC CTGGAGTGAT TTTAACGCCT TACACGTTGA CCTGTCCGTC TAATACATCC
    c-ukb22 AATTTGTGCC CTCGAGTGAT TTCAACGCCT TACACGTTGA CCTGTCCGTC TAATACATCC
     c-gk18 AATCTCTGCC CTGGAGTGAT TTCAACGCCT TACACGTTGA CCTGTCCGTC TAATACATCC
  c-k1-8-dem AATTTGTGCC CTGGAGTGAT TTCAACGCCT TACACGTTGA CCTGTCCGTC TGATGCAACC
    c-ugd23 AATTTGTGCC CTGGAGTGAT TTCAACGCCT TACACGTTGA CCTGTCCGTC TGATGCAACC
    a-ugd16 AATTTGTGCC CTGGAGTGAT TTCAACGCCT TACACGTTGA CCTGTCTGTC TGATGCATCC
    a-ugd18 AATTTGTGCC CTGGAGTGAT TTCAACGCCT TACACGTTGA CCTGTCTGTC TGATGCATCC
    a-ugd24 AATTTGTGCC CTGGAGTGAT TTCAACGCCT TACACGTTGA CCTGTCTGTC TGATGCATCC
    a-ug374 AATTTGTGCC CTGGAGTGAT TTCAACGCCT TACACGTTGA CCTGTCTGTC TGATGCATCC
     a-ugd4 AATTTGTGCC CTGGAGTGAT TTCAACGCCT TACACGTTGA CCTGTCTGTC TGATGCATCC
    a-bcbl1 AATTTGTGCC CTGGAGTGAT TTCAACGCCT TACACGTTGA CCTGTCTGTC GGATTCATCC
    a-uka13 AATTTGTGCC CTGGAGTGAT TTCAACGCCT TACAATTTGA CCTGTCTGTC TAATGCATCC
    a-uka21 AATTTGTGCC CTGGAGTGAT TTCAACGCCT TACAATTTGA CCTGTCTGTC TAATGCATCC
    a-ukma4 AATTTGTGCC CTGGAGTGAT TTCAACGCCT TACAATTTGA CCTGTCTGTC TAATGCATCC
      a-bc1 AATTTGTGCC CTGGAGTGAT TTCAACGCCT TACACGTTGA CCTGTCTGTC TAATGCATCC
     a-ema7 AATTTGTGCC CTGGAGTGAT TTCAACGCCT TACAACTTGA CCTGTCTGTC TAATGCATCC
    a-bcblr AATTTGTGCC CTGGAGTGAT TTCAACGCCT TACAAGTTGA CCTGTCTGTC TAATGCATCC
    a-ukb12 AATTTGTGCC CTGGAGTGAT TTCAACGCCT TACACGTTGA CCTGTCTGTC TAATGCATCC
    a-ug111 AATTTGTGCC CTGGAGTGAT TTCAACGCCT TACAATTTGA CCTGTCTGTC TAATGCATCC
     a-iap1 AATTTGTGCC CTGGAGTGAT TTCAACGCCT TACACGTTGA CCTGTCTGTC TGATACATCC
    a-ukc12 AATTTGTGCC CTGGAGTGAT TTCAACGCCT TACAAGTTGA CCTGTCTGTC TAATGCATCC
c-k1-43-ber CATTTGTGCC CTGGAGTGAT TTCAACGCCT TACACGTTGA CCTGTCCGTC TGATGCATCC
    a-ugd12 AATTTGTGCC CTGGAGTGAT TTCAACGCCT TACACGTTGA CCTGTCTGTC TGATGCATCC
  Consensus ----T-TGCC CT--AGTG-- TT--ACG--T TACA--TTGA C-TGTC--TC ------A--C
```

```
            121                                                                 180
   d1-tks10 TTGCCAATAT CCTGGTATTG CAACGGAACT CGGCTTCTTC GAATTACTGG GGCAACACTG
    d2-zks3 TTGCCAATAT CCTGGTATTG CAACGGAACT TTGCTTATGC GATATCACAG GACCACACTA
     b-ugd1 TTGCCAACAT CCTGGTATTG CAACGGGACT CAGCTTCGGC GAATACGGGG GTCTAACCTA
    b-ugd26 TTGCCAATAT CCTGGTATTG CAACGGGACT CAGCTTCGGC GAATAAGGGA GTCTACCCTA
    b-ugd19 TTGCCAATAT CCTGGTATTG CAACGGGACT CGGCTTTGGC GAATAAGGGA GTCTAACCTA
     b-ugd7 TTGCCAATAT CCTGGTATTG CAACGGGACT CAGCTTTGGC GAATAACGGA CTCTACCCTA
     b-ugd2 TTGCCAATAT CCTGGTATTG CAACGGGACT CAGCTTTCGC GAATAAGGGC GTCTACCCTA
    b-ugd21 TTGCCAATAT CCTGGTATTG CAACGGGACT CAGCTTTGGC GAATAAGGGA GAATACCCTA
    b-ugd10 TTGCCAATAT CCTGGTATTG CAACGGGACT CAGCTTCACC GAATAACGGC GTCTAACCTA
   b-ukma24 TTGCCAATAT CCTGGTATTG CAACGGGACT CGGCTTCACC GACTAACGGC GTCTAACCTA
    b-ugd29 TTGCCAATAT CCTGGTATTG CAACGGGACT CGGCTTCACC GAATAACGGC GTCTAACCTA
     b-ug52 TTGTCAATAT CCTGGTATTG CAACGGGACT CGGCTTCACC GAATAACGGC GTCTAACCTA
     b-g413 TTGCCAATAT CCTGGTATTG CAACGGGACT CGGCTTTTGC GAATAACGGC GTCTAACCCA
     b-g482 TTGCCAATAT CCTGGTATTG CAACGGGACT CGGCTTTGGC GAATAACGGC GTCTAACCTA
    b-ugd13 TTGCCAATAT CCTGGTATTG CAACGGGACT CGGCTTTGGC GAATAACGGC GTCTAACCTA
     b-ugd3 TTGCCAATAT CCTGGTATTG CAACGGGACT CGGCTTTGGC GAATAACGGC GTCTAACCTA
    b-ugd15 TTGCCAATAT CCTGGTATTG CAACGGGACT CGGCTTTGGC GAATAACGGC GTCTAACCTA
     b-g71 TTGCCAATAT CCTGGTATTG CAACGGGACT CGGCTTTGGC GAATAACGGC GTCTAACCTA
     b-g91 TTGCCAATAT CCTGGTATTG CAACGGGACT CAGCTTTTGC GAATAACGGC GTCTAACCTA
     c-gk17 TTGCCAACAT CCTGGTATTG CAACGATACT CGGCTTTTCC GACTGACGCA GGACACATTC
     c-iap3 TTGCCAACAT CCTGGTATTG CAACGATACT CGGCTTTTCC GACTGACGCA GGACACATTC
     c-erla TTGCCAACAT CCTGGTATTG CAACGATACT CGGCTTCTCC GACTGACGCA GCAAACATTC
    c-icam1 TTGCCAACAT CCTGGTATTG CAACGATACT CGGCTTCTCC GACTGACGCA GCAAACATTC
    c-ukma1 TTGCCAACAT CCTGGTATTG CAACGATACT CGGCTTTTCC GACTGACGCA GCAAACATGG
     c-iap2 TTGCCAACAT CCTGGTATTG CAACGATACT CGGCTTTTCC GACTGACGCA GCAAACATTT
     c-bbg1 TTGCCAACAT CCTGGTATTG CAACGATACT CGGCTTCTCC GACTGACGCA GCAAACATTC
    c-ukma8 TTGCCAACAT CCTGGTATTG CAACGATACT CGGCTTCTCC GACTGACGCA GCAAACATTC
      c-bc2 TTGCCAACAT CCTGGTATTG CAACGATACT CGGCTTCTCC GACTGACGCA GCAAACATTC
    c-ukma3 TTGCCAACAT CCTGGTATTG CAACGATACT CGGCTTCTCC GACTGACGCA GCAAACAATC
     c-ive1 TTGCCAACAT CCTGGTATTG CAACGATACT CGGCTTTTCC GACTGACGCA GCAAACATTA
    c-ukb22 TTGCCAACAT CCTGGTATTG CAACGATACT CGGCTTTTCC GACTGACGCA GCAAACATTG
     c-gk18 TTGCCAACAT CCTGGTATTG CAACGATACT CGGCTTTTAC GAGTGACGCA GGGAACATTG
  c-k1-8-dem TTGCCAATAT CCTGGTATTG CAACGATACT CGGCTTTTCC GACTGACGCA GCAAACACTC
    c-ugd23 TTGCCAATAT CCTGGTATTG CAACGATACT CGGCTTCTCC GACTGACGCA GCAAACATTC
    a-ugd16 TTGCCAATAT CCTGGTATTG CAACGATACT CGGCTTTTGC GACTGACGGA CCAATCATTC
    a-ugd18 TTGCCAATAT CCTGGTATTG CAACGATACT CGGCTTTTGC GACTGACGGA CCAATCATTC
    a-ugd24 TTGCCAATAT CCTGGTATTG CAACGATACT CGGCTTTTGC GACTGACGGA CCAATCATTC
    a-ug374 TTGCCAATAT CCTGGTATTG CAACGATACT CGGCTTTGGC GACTGACGGA CCAATCATTC
     a-ugd4 TTGCCAATAT CCTGGTATTG CAACGATACT CGGCTTTGGC GACTGACGGA CCAATCATTC
    a-bcbl1 TTGCCAATAT CCTGGTATTG CAACGATACT CGGCTTTGGC GACTGACGAA GCCAACACTC
    a-uka13 TTGCCAATCT CCTGGTATTG CAACAATACT CGGCTTTTCC GACTGACGGA GAGAACATTG
    a-uka21 TTGCCAATCT CCTGGTATTG CAACAATACT CGGCTTTTCC GACTGACGGA GAGAACATTG
    a-ukma4 TTGCCAATCT CCTGGTATTG CAACAATACT CGGCTTTTAC GACTGACGGA GAGAACATTG
      a-bc1 TTGCCAATAT CCTGGTATTG CAACAATACT CGGCTTTTGC GACTGACGGA GAGAAGAGTC
     a-ema7 TTGCCAATAT CCTGGTATTG CAACAATACT CGGCTTTTGC GACTGACGGA GAGAAGAGTC
    a-bcblr TTGCCAATAT CCTGGTATTG CAACAATACT CGGCTTTTCC GACCGACGGA GACAACACTT
    a-ukb12 TTGCCAATAT TCTGGTATTG CAACAATACT CGGCTCTTCC GACTGACGAA GACAATATTT
    a-ug111 TTGCCAATAT CCTGGTATTG CAACAATACT CGGCTTTTCC GACTGACGGA GAAAACACTC
     a-iap1 TTGCCAATAT CCTGGTATTG CAACAATACT CGGCTTTTGC GACTGACGAC GAAAACACTC
    a-ukc12 TTGCCAATAT CCTGGTATTG CAACGATACT CGGCTTTTCC GACTGACGGA GAGAACACTT
c-k1-43-ber TTGCCAATAT CCTGGTATTG CAACGGAACT CAGCTTCTTC GACTGACGCA GCGATCAGTA
    a-ugd12 TTGCCAATAT CCTGGTATTG CAACGATACT CGGCTTTGGC GACTGACGGA CCAATCATTC
  Consensus TTG-CAA--T -CTGGTATTG CAAC---ACT --GCT----C GA-------- ----------
```

```
            181                                                                240
   d1-tks10 ACTATTCCTT CCCTTACCGG CAATTTTACT TGTGTGGATC ACTCTGGCCT TTCACACAGC
    d2-zks3 ACTCTTATGA ACCTTGCCGC CAATTGGACT TGTGTGAATC AATCTGGAAT TTCACACAGC
     b-ugd1 ACTGTTTCTT TGCTCACCTG CAATTTTACT TGTATGACAG CATCTGGGCC TACACACAGC
    b-ugd26 ACTGTTTCTT TCCTCACCTG CAATTTTACT TGTATGACAG CATCTGGGCC TACACACAGC
    b-ugd19 ACTGTTGCTT CGCTCACCGG CAATTTTACT TGTATGACAG CATCTGGGCC TACACACAGC
     b-ugd7 ACTGTTTCTT CGCTCACCGG CAATTTTACT TGTATGACAG CATCTGGGCC TACACACAGC
     b-ugd2 ATTGTTTCTT CGCTCACCGG CAATTTTACT TGTATGACAG CATCTGGGCC TACACACAGC
    b-ugd21 ACTGTTTCTT CGCTCACCGG CAATTTTACT TGTATGACAG CATCTGGGCC TACATACAGC
    b-ugd10 ACTGTTTCTT CGCTCACCTG CAATTTTACT TGTATGACAA CATCTGGGCC TACACACAGC
   b-ukma24 ACTGTTTCTT CGCTCACCTG CAATTTTACT TGTATGACAA CATCTGGGCC TACACACAGC
    b-ugd29 ACTGTTTCTT CGCTCACCTG CAATTTTACT TGTATGACAA CATCTGGGCC TACACACAGC
     b-ug52 ACTGTTTCTT CGCTCACCTG CAATTTTACT TGTATGACAA CATCTGGGCC TACACACAGC
     b-g413 ACTGTTTGTT CGTTAACCTG CAATTTTACT TGTATGACAG CATCTGGGCC TACACACAGC
     b-g482 ACTGTTTCTC GGTTAACCTG CAATTTTACT TGTATGACAA GATCTGGGCC TACACACAGC
    b-ugd13 ACTGTTTTGT CGGTGACCTG CAATTTTACT TGTATGACAA CATCTGGGCC TACACACAGC
     b-ugd3 ACTGTTTTGT CGGTGACCTG CAATTTTACT TGTATGACAA CATCTGGGCC TACACACAGC
    b-ugd15 ACTGTTTTGT CGGTGACCTG CAATTTTACT TGTATGACAA CATCTGGGCC TACACACAGC
     b-g71 ACTGTTTCTT CGCTCAGCTG CAATTTTACT TGTATGACAA CATCTGGGCC TACACACAGC
     b-g91 ACTGTTTCTG CGATGACCTG CAATTTTACT TGTATGACAA CATCTGGGCC TACACACAGC
     c-gk17 ACTGTTGTCA ACTTTATGTG CAATTTTTCT TGTGTGGGAC AATCTGGGCA TCGACACAGC
     c-iap3 ACTGTTGTCA ACCTTATGTG CAATTTTTCT TGTGTGGGAC AATCTGGGCA TCGACACAGC
     c-erla ACTGTTGACA CCCTTATCTG CAATTTTAGT TGTGTGGGAC AATCTGGGCA TCGACACAGC
    c-icam1 ACTGTTGTCG CCCTTATCTG CAATTTTAGT TGTGTGGGAC AATCTGGGCA TCGACACAGC
    c-ukma1 ACTGTTCACA CCCTTATCTG CAATTTTAGT TGTGTGGGAC AATCTGGGCA TCGACACAGC
     c-iap2 ACTGTTAACA CCCTTATCTG CAATTTTACT TGTGTGGGAC AATCTGGGCA TCGACACAGC
     c-bbg1 ACTGTTGTCA CCCTTATCTG CAATTTTAGT TGTGTGGAAC AATCTGGGCA TCGACACAGC
    c-ukma8 ACTGTTGTCA CCCTTATCTG CAATTTTAGT TGTGTGGGAC AATCTGGGCA TCGACACAGC
      c-bc2 ACTGTTGTCG CCCTTATCTG CAATTTTAGT TGTGTGGGAC AATCTGGGCA TCGACACAGC
    c-ukma3 ACTGTTGTCA CCCTTATCTG CAATTTTAGT TGTGTGGGAC AATCTGGGCA TCGACACAGC
     c-ive1 ACTGTTGTCA CCCTTATCTG CAATTTTAGT TGTGTGGGAC AATCTGGGCA TCGACACAGC
    c-ukb22 ACTGTTGTTG ACCTTATCTG CAATTTTAGT TGTGTGGGAC AATCTGGGCA TCGACACAGC
     c-gk18 ACTGTTGACA CCCTTATCTG CAATTTTAGT TGTGTGGGAC AATCTGGGCA TCGATACAGC
  c-k1-8-dem ACTTTTACCA ACCTTACCTG CAATTTTACT TGTGTGGGAC AATCTGGGCA TCGACACAGC
    c-ugd23 ACTTTTGCCA ACCTTACCTG CAATTTTACT TGTGTGGGAC AATCTGGGTA TCGACACAGC
    a-ugd16 ACTGTTGCCA CCATTACCTG CAATTTTACT TGTGTGGAAC AATCTGGGCA TCGACAGAGC
    a-ugd18 ACTGTTGCCA CCATTACCTG CAATTTTACT TGTGTGGAAC AATCTGGGCA TCGACAGAGC
    a-ugd24 ACTGTTGCCA CCATTACCTG CAATTTTACT TGTGTGGAAC AATCTGGGCA TCGACAGAGC
    a-ug374 ACTGTTGCCA CCATTACCTG CAATTTTACT TGTGTGGAAC AATCTGGGCA TCGACAGAGC
     a-ugd4 ACTGTTGCCA CCATTACCTG CAATTTTACT TGTGTGGAAC AATCTGGGCA TCGACAGAGC
    a-bcbl1 ACTATTGACA TCATTACCTG CAATTTTACT TGTGTGGAAC AATCTGGGCA TCGACAGAGC
    a-uka13 TTTTCTGTCA CCATTGCCTG CAATTTTACT TGTGTGGAAC AATCTGGGCA TCGACAGAGC
    a-uka21 TTTTCTGTCA CCATTGCCTG CAATTTTACT TGTGTGGAAC AATCTGGGCA TCGACAGAGC
    a-ukma4 TTTTCTGTCA CCATTGCCTG CAATTTTACT TGTGTGGAAC AATCTGGGCA TCGACAGAGC
      a-bc1 ATTCTTGACA CCATTGCCTG CAATTTTACT TGTGTGGAAC AATCTGGGCA TCGACAGAGC
     a-ema7 ATCCTTGTCA CCATTGCCTG CAATTTTACT TGTGTGGAAC AATCTGGGCA TCGACAGAGC
    a-bcblr TTTCCTGTCA CCATTGCCTG CAATTTTACT TGTGTGGAAC AATCTGGGCA TCGACAGAGC
    a-ukb12 TCTACTGTCA CCATTGCCTG CAATTTTACT TGTGTGGAAC AATCTGGGCA TCGACAGAGC
    a-ug111 CCTGGTGTCA ACATTGCCTG CAATTTTACT TGTGTGGAAG AATCTGGGCA TCCACAGAAC
     a-iap1 TATACTTTCA CCATGACCTG CAATTTCACT TGTGTGGAAC AATCTGGGCA TCGACAGAGC
    a-ukc12 TTTCCTGTCA CCATTCCCTG CAATTTTACT TGTGTGGAAC AATCTGGGCA TCGACAGAGC
c-k1-43-ber CCTGTTTCCA CACTTACCTG CAATTTTACT TGTGTGGGAC AATCTGGGCC TTCACACAGC
    a-ugd12 ACTGTTGCCA ACATTACCTG CAATTTTACT TGTGTGGAAC AATCTGGGCA TCGACAGAGC
  Consensus -----T---- ---T------ CAATT----T TGT-TG---- --TCTGG--- T--A-A-A-C
```

```
            241                                                        300
   d1-tks10 ATTTGGATTC AACGGTATCC ACCACCTGTC TTACAAACCT TGTGTGCACA GCCATCAACC
    d2-zks3 ATTTGGATTC AATGGTATAC AGAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
     b-ugd1 ATTTGGATTG AATGGTATAC AACACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
    b-ugd26 ATTTGGATTG AATGGTATAC AACACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
    b-ugd19 ATTTGGATTG AATGGTATAC AACACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
     b-ugd7 ATTTGGATTG AATGGTATAC AACACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
     b-ugd2 ATTTGGATTG AATGGCATAC AACACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
    b-ugd21 ATTTGGATTG AATGGTATCC AACACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
    b-ugd10 ATTTGGATTC AATGGTATAC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
   b-ukma24 ATTTGGATTG AATGGTATAC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
    b-ugd29 ATTTGGATTG AATGGTATAC ACAACCTGTC TTACAAACCT TGTGTACACA GCCATCAAAC
     b-ug52 ATTTGGATTG AATGGTATAC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
     b-g413 ATTTGGATTG AATGGTATAC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
     b-g482 ATTTGGATTG AATGGTATAC ACAACCTGTC TTACAAACCT CGTGTGCACA GCCATCAAAC
    b-ugd13 ATTTGGATTG AATGGTATTC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
     b-ugd3 ATTTGGATTG AATGGTATTC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
    b-ugd15 ATTTGGATTG AATGGTATTC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
     b-g71 ATTTGGATTG AATGGTATAC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
     b-g91 ATTTGGATTG AATGGTATAC AGAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
     c-gk17 CTTTGGATGG CATGGTATGC ACCACCTGTC TTACAAACCT TGTGTGGACA ACCAGCAAAC
     c-iap3 CTTTGGATGG CATGGTATGC ACCACCTGTC TTACAAACCT TGTGTGGACA ACCAGCAAAC
     c-erla CTTTGGATTA CATGGTATCC ACAACCTGTC TTACAAACCT TGTGTGGACA ACCATCAAAC
    c-icam1 CTTTGGATTA CATGGTATCG AGAACCTGTC TTACAAACCT TGTGTGGACA ACCATCAAAC
    c-ukma1 CTTTGGATTA CATGGTATCC ACAACCTGTC TTACAAACCT TGTGTGGACA ACCATCAAAC
     c-iap2 CTTTGGATTA CATGGTATGC ACAACCTGTC TTACAAACCT TCTGTGGACA ACCATCAAAC
     c-bbg1 CTTTGGATTA CATGGTATCC ACAACCTGTC TTACAAACCT TGTGTGGACA ACCATCAAAC
    c-ukma8 CTTTGGATTA CATGGTATCC ACAACCTGTC TTACAAACCT TGTGTGGACA ACCATCAAAC
      c-bc2 CTTTGGATTA CATGGTATCC ACAACCTGTC TTACAAACCT TGTGTGGACA ACCATCAAAC
    c-ukma3 CTTTGGATTA CATGGTATCC ACAACCTGTC TTACAAACCT TGTGTGGACA ACCATCAAAC
     c-ive1 CTTTGGATTA CATGGTATCC ACAACCTGTC TTACAAACCT TGTGTGGACA ACCATCAAAC
    c-ukb22 CTTTGGATTA CATGGTATCC ACAACCTGTC TTACAAACCT TGTGTGGACA ATCATCAAAC
     c-gk18 CTTTGGATTA CATGGTATGC ACAACCTGTC TTACAAACCT TCTGTGGACA ACCATCAAAC
  c-k1-8-dem TTTTGGATTA CATGGCATGC ACAACCTGTC TTACAAACCT TGTGTGCACA ACCATCAAAC
    c-ugd23 GTTTGGATTA CATGGCATGC ACAACCTGTC TTACAAACCT TGTGTGCGCA ACCATCAAAC
    a-ugd16 ATTTGGATTA CATGGAATGC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
    a-ugd18 ATTTGGATTA CATGGAATGC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
    a-ugd24 ATTTGGATTA CATGGAATGC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
    a-ug374 ATTTGGATTA CATGGAATGC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
     a-ugd4 ATTTGGATTA CATGGAATGC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
    a-bcbl1 ATTTGGATTA CATGGCATGC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCTAAC
    a-uka13 ATTTGGATTA CATGGCATGC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
    a-uka21 ATTTGGATTA CATGGCATGC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
    a-ukma4 ATTTGGATTA CATGGCATGC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
      a-bc1 ATTTGGATTA CATGGCGTGC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
     a-ema7 ATTTGGATTA CATGGCATGC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
    a-bcblr ATTTGGATTA CATGGCATGC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
    a-ukb12 ATTTGGATTA CATGGCATGC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
    a-ug111 ATTTGGATTA CATGGCATGC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
     a-iap1 ATTTGGATTT CATGGCATAC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
    a-ukc12 ATTTGGGTTA CATGGCATGC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
c-k1-43-ber ATTTGGATCA CATGGTATGC AAAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
    a-ugd12 ATTTGGATTA CATGGAATGC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
  Consensus -TTTGG-T-- -A-GG--T-- A--ACCTGTC TTACAAACCT --TGT---CA --CA-C-A-C
```

```
              301                                                       360
   d1-tks10   ACAGTTACTT GTGGTCAGCG TGTTAGTTTG CATTGTTCTA CCTCTGCAAA TACTGTTACC
    d2-zks3   ACAGTTACTT GTGGTCAGCG TGTTACTTTG CATTGCTCTA CCTCTGCAAG TACTGTTTTT
     b-ugd1   ACAGTGACTT GTGGTCAGCG TGTTACTTTG TATTGTCATA CCTCTTCAAA TAATGGTACC
    b-ugd26   ACAGTGACTT GTGGTCAGCG TGTTACTTTG TATTGTTATA CCTCTTCAAA TAATGGTACC
    b-ugd19   ACAGTGACTT GTGGTCAGCG TGTTACTTTG TATTGTCATA CCTCTTCAAA TAATGGTACC
     b-ugd7   ACAGTGACTT GTGGTCAGCG TGTTACTTTG TATTGTCATA CCTCTTCAAA TAATGGTACC
     b-ugd2   ACAGTGACTT GTGGTCAGCG TGTTACTTTG TATTGTCATA CCTCTTCAAA TAATGGTACC
    b-ugd21   ACAGTGACTT GTGGTCAGCG TGTTACTTTG TATTGTCATA CCTCTTCAAA TAATGGTACC
    b-ugd10   ACAGTGTCTT GTGGTCAGCC TGTTACTTTG TATTGTGATA CCTCTTCAAA TAATGTTACC
   b-ukma24   ACAGTGTCTT GTGGTCAGCC TGTTACTTTG TATTGTGATA CCTCTTCAAA TAATGTTACC
    b-ugd29   ACAGTGTCTT GTGGTCAGCC TGTTACTTTG TATTGTGATA CCTCTTCAAA TAATGTTACC
     b-ug52   ACAGTGTCTT GTGGTCAGCC TGTTACTTTG TATTGTGATA CCTCTTCAAA TAATGTTACC
     b-g413   ACAGTGACTT GCGGTCAGCG TGTTACTTTG TATTGTCATA CCTCTTCAAA TAATGTTACC
     b-g482   ACAGTGACTT GCGGTCAGCG TGTTACTTCG TATTGTCATA CCTCTTCAAA TAATGTTAAC
    b-ugd13   ACAGTGACTT GCGGTCAGCG TGTTACTTTG TATTGTCATA CCTCTTCAGA TAATGTTACC
     b-ugd3   ACAGTGACTT GCGGTCAGCG TGTTACTTTG TATTGTCATA CCTCTTCAGA TAATGTTACC
    b-ugd15   ACAGTGACTT GCGGTCAGCG TGTTACTTTG TATTGTCATA CCTCTTCAGA TAATGTTACC
     b-g71    ACAGTGACTT GCGGTCAGCG TGTTACTTTG TATTGTCATA CCTCTTCAAA TAATGTTACC
     b-g91    ACAGTGACTT GCGGTCAGCG TGTTACTTTG TATTGTAATA CCTCTTCAAA TAATGTTACC
     c-gk17   ACAGTCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
     c-iap3   ACAGTCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
     c-erla   ACAGTCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
    c-icam1   ACAGTCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
    c-ukma1   ACAGTCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
     c-iap2   ACAGTCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
     c-bbg1   ACAGTCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
    c-ukma8   ACAGTCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
      c-bc2   ACAGTCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
    c-ukma3   ACAGTCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
     c-ive1   ACAGTCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
    c-ukb22   ACAGTCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
     c-gk18   ACAGTCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
  c-k1-8-dem  ACAGTCACTT GTGCTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
    c-ugd23   ACAGTCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
    a-ugd16   ACAGTCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
    a-ugd18   ACAGTCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
    a-ugd24   ACAGTCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
    a-ug374   ACAGTCGCTT GTGGTCAGCA GGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
     a-ugd4   ACAATCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
    a-bcbl1   ACAGTCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
    a-uka13   ACAGTCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
    a-uka21   ACAGTCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
    a-ukma4   ACAGTCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
      a-bc1   ACAGTCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
     a-ema7   ACAGTCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
    a-bcblr   ACAGTCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
    a-ukb12   ACAGTCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
    a-ug111   ACAGTCACTT GTGGTCACCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
     a-iap1   ACAGTCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
    a-ukc12   ACAGTCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
c-k1-43-ber   ACAGTCACTT GTGGTCAGCA TGTTCCTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
    a-ugd12   ACAGTCGCTT GTGGTCAGCA GGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
  Consensus   ACA-T--CTT G-G-TCA-C- -GTT--TT-G -ATTG---TA CCTCT--A-- TA-TG-T---
```

```
            361                                                        420
   d1-tks10 ATTTGGCGTC TACAAAACGA AGGACATCAA CCCGTGTCAC AAACTTACTA CTATAATTTT
    d2-zks3 ATTTGGCGTC TACACAGGGG AGGAAATCAA ACCGTGTCAC AAACTAACTA CTATAATTTT
     b-ugd1 ATTTGGCGTC TACAAAACGG ACACAATCAA ACCGTGTCAC AAACTAAATA CTATAATTTT
    b-ugd26 ATTTGGCGTC TACAAAACGG ACACAATCAA ACCGTGTCAC AAACTAAATA CTATAATTTT
    b-ugd19 ATTTGGCGTC TACAAAAGGG ACACAATCAA ACCGTGTCAC AAACTAAATA CTATAATTTT
     b-ugd7 ATTTGGCGTC TACAAAACGG ACACAATCAA ACCGTGTCAC AAACTAAATA CTATAATTTT
     b-ugd2 ATTTGGCGTC TACAAAACGG ACACAATCAA ACCGTGTCAC AAACTAAATA CTATAATTTT
    b-ugd21 ATTTGGCGTC TACAAAACGG ACACAATCAA ACCGTGTCAC AAACTAAATA CTATAATTTT
    b-ugd10 ATTTGGCGTC TACAAAACGG ACACAATCAA ACCGTGTCAC AAACTAAATA CTATAATTTT
   b-ukma24 ATTTGGCGTC TACAAAACGG ACACAATCAA ACCGTGTCAC AAACTAAATA CTATAATTTT
    b-ugd29 ATTTGGCGTC TACAAACCGG ACACAATCAA ACCGTGTCAC AAACTAAATA CTATAATTTT
     b-ug52 ATTTGGCGTC TACAAAACGG ACACAATCAA ACCGTGTCAC AAACTAGATA CTATAATTTT
     b-g413 ATTTGGCGTC TACAAAACGG ACACAATCAA ACCGTGTCAC AAACTAAATA CTATAATTTT
     b-g482 ATTTGGCGTC TACAAAACGG ACACAATCAA ACCGTGTCAC AAACTAAATA CTATAATTTT
    b-ugd13 ATTTGGCGTC TACAAAACGG ACACAATCAA ACCGTGTCAC AAACTAAATA CTATAATTTT
     b-ugd3 ATTTGGCGTC TACAAAACGG ACACAATCAA ACCGTGTCAC AAACTAAATA CTATAATTTT
    b-ugd15 ATTTGGCGTC TACAAAACGG ACACAATCAA ACCGTGTCAC AAACTAAATA CTATAATTTT
     b-g71 ATTTGGCGTC TACAAAACGG ACACAATCAA ACCGTGTCAC AAACTAAATA CTATAATTTT
     b-g91 ATTTGGCGTC TACAAAACGG ACACAATCAA ACCGTGTCAC AAACTAAATA CTATAATTTT
     c-gk17 GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
     c-iap3 GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
     c-erla GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
    c-icam1 GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
    c-ukma1 GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
     c-iap2 GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
     c-bbg1 GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
    c-ukma8 GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
      c-bc2 GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
    c-ukma3 GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
     c-ive1 GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
    c-ukb22 GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
     c-gk18 GTTTGGCATC TACCAAACGG ACAAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
  c-k1-8-dem GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
    c-ugd23 GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
    a-ugd16 GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
    a-ugd18 GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
    a-ugd24 GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
    a-ug374 GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
     a-ugd4 GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
    a-bcbl1 GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
    a-uka13 GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
    a-uka21 GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
    a-ukma4 GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
      a-bc1 GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
     a-ema7 GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
    a-bcblr GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
    a-ukb12 GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
    a-ug111 GTTTGGCATC TACCAAACGG ACTAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
     a-iap1 GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
    a-ukc12 GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
c-k1-43-ber ATTTGGCATC TACAAAACGG ACGAAATCAA ACCGTGTCAC AAACTAACTA CTATAATTTT
    a-ugd12 GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
  Consensus -TTTGGC-TC TAC--A--G- A----AT-AA -CCGTGTCAC AAACT---TA CTATAATTTT
```

```
            421                                                        480
   d1-tks10 ACGCTGATGA ACCAAACTCA GGGATGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCATG
    d2-zks3 ACGCTGATGA ACCAAACTAA CGGCTGTTAT GCTTGTTCTA GCGGGCTGTC GTCTCGCCTG
     b-ugd1 ACGATGATGA ACCAAACTCA GGGGTGTTAT CTTTGTTCTG ACGGGCTGTC GTCTCGCCTG
    b-ugd26 ACGATGATGA ACCAAACTCA GGGGTGTTAT CTTTGTTCTG ACGGGCTGTC GTCTCGCCTG
    b-ugd19 ACGATGATGA ACCAAACTGA GGGGTGTTAT CTTTGTTCTG ACGGGCTGTC GTCTCGCCTG
     b-ugd7 ACGATGATGA ACCAAACTCA GGGGTGTTAT TTTTGTTTTG ACGGGCTGTC GTCTCGCCTG
     b-ugd2 ACGATGATGA ACCAAACTCA GGGGTGTTAT ATTTGTTGTG ACGGGCTGTC GTCTCACCTG
    b-ugd21 ACGATGATGA ACCAAACTCA GGGGTGTTAT ATTTGTTGTG ACGGGCTGTC GTCTCGCCTG
    b-ugd10 ACGATGATGA ACCAAACTCA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
   b-ukma24 ACGATGATGA ACCAAACTCA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
    b-ugd29 ACGATGATGA ACCAAACTCA GGGGTGTTAT GCTTGTTCTA ACGGGCTGTC GTCTCGCCTG
     b-ug52 ACGATGATGA ACCAAACTCG GGGGTGTTAT GTTTGTTCTG ACGGGCTGTC GTCTCGCCTG
     b-g413 ACGATGATGA ACCAAACTCA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
     b-g482 ACGATGATGA ACCAAACTCA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
    b-ugd13 ACGATGATGG ACCAAACTCA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
     b-ugd3 ACGATGATGG ACCAAACTCA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
    b-ugd15 ACGATGATGA ACCAAACTCA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
     b-g71  ACGATGATGA ACCAAACTCA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
     b-g91  ACGATGATGA ACCAAACTCA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
     c-gk17 ACGCTGATGA ACCAAACTGA GGGGTGTTAT GCTTGTTCTA ACGGGCTGTC GTCTGGCCTG
     c-iap3 ACGCTGATGA ACCAAACTGA GGGGTGTTAT GCTTGTTCTA ACGGGCTGTC GTCTGGCCTG
     c-erla ACGCTGATGA ACCAAACTGA GGGGTGTTAT GCTTGTTCTA ACGGGCTGTC GTCTCGCCTG
    c-icam1 ACGCTGATGA ACCAAACTGA GGGGTGTTAT GCTTGTTCTA ACGGGCTGTC GTCTCGCCTG
    c-ukma1 ACGCTGATGA ACCAAACTGA GGGGTGTTAT GCTTGTTCTA ACGGGCTGTC GTCTCGCCTG
     c-iap2 ACGCTGATGA ACCAAACTGA GGGGTGTTAT GCTTGTTCTA ACGGGCTGTC GTCTCGCCTG
     c-bbg1 ACGCTGATGG ACCAAACGGA GGGGTGTTAT GCTTGTTCTG ACGGGCTGTC GTCTCGCCTG
    c-ukma8 ACGCTGATGG ACCAAACGGA GGGGTGTTAT GCTTGTTCTG ACGGGCTGTC GTCTCGCCTG
      c-bc2 ACGCTGATGG ACCAAAGTGA GGGGTGTTAT GCTTGTTCTG ACGGGCTGTC GTCTCGCCTG
    c-ukma3 ACGCTGATGG ACCAAACGGA GGGGTGTTAT GCTTGTTCTG ACGGGCTGTC GTCTCGCCTG
     c-ive1 ACGCTGATGG ACCAAACGGA GGGGTGTTAT GCTTGTTCTG ACGGGCTGTC GTCTCGCCTG
    c-ukb22 ACGCTGATGG ACCAAACGGA GGGGTGTTAT GCTTGTTCTG ACGGGCTGTC GTCTCGCCTG
     c-gk18 ACGCTCATGA ACCAAACTGA GGGGTGTTAT GCTTGTTCTA ACGGGCTGTC GTCTCGCCTG
  c-k1-8-dem ACGCTGATGA ACCAAACTGA GGGGTGTTAT GCTTGTTCTA ACGGCCTGTC GTCTCGCCTG
    c-ugd23 ACGCTGATGA ACCAAACTGA GGGGTGTTAT GGTTGTTCTA ACGGCCTGTC GTCTCGCCTG
    a-ugd16 ACGCTGATGA ACCAAACTGA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
    a-ugd18 ACGCTGATGA ACCAAACTGA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
    a-ugd24 ACGCTGATGA ACCAAACTGA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
    a-ug374 ACGCTGATGA ACCAAACTGA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
     a-ugd4 ACGCTGATGA ACCAAACTGA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
    a-bcbl1 ACGCTGATGA ACCAAACTGA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
    a-uka13 ACGCTGATGA ACCAAACTGA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
    a-uka21 ACGCTGATGA ACCAAACTGA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
    a-ukma4 ACGCTGATGA ACCAAACTGA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
      a-bc1 ACGCTGATGA GCCAAACTGA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
     a-ema7 ACGCTGATGA ACCAAACTGA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
    a-bcblr ACGCTGATGA ACCAAACTGA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
    a-ukb12 ACGCTGATGA ACCAAACTGA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
    a-ug111 ACGCCGATGA ACCAAACTGA AGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
     a-iap1 ACGCTGATGA ACCAAACTGA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
    a-ukc12 ACGCTGATGA ACCAAACTGA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
c-k1-43-ber ACGGTGATGA ACCAAACTGA GGGGTGTTAT GCTTGTTCTA ACGGGCTGTC GTCTCGCCTG
    a-ugd12 ACGCTGATGA ACCAAACTGA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
  Consensus ACG---ATG- -CCAAA---- -GG-TGTTAT --TTGTT-T- -CGG-CTGTC GTCT--C-TG
```

```
            481                                                        540
   d1-tks10 TCAAATCGTA TATGTTTTTG GGCGCCTTGT GCCAATATAA CTCCAGAAAC TGATACTGTA
    d2-zks3 TCAAATCTTA TATGTTTTTG GGCGCATTGT GCCAATATAT CTCTAGAAAC TTCTACTGTA
     b-ugd1 TCAAATCGTA TATGTTTTTG GGCGTCTTGT ATCAATATAA CTCCAGAAAC TCCTACTGTC
    b-ugd26 TCAAATCGTA TATGTTTTTG GGCGTCTTGT ATCAATATAA CTCCAGAAAC TCCTACTGTC
    b-ugd19 TCAAATCGTA TATGTTTTTG GGCGTCTTGT ATCAATATAA CTCCAGAAAC TCCTACTGTC
     b-ugd7 TCAAATCGTA TATGTTTTTG GGCGTCTTGT ATCAATATAA CTCCAGAAAC TCCTACTGTC
     b-ugd2 TCAAATCGTA TATGTTTTTG GGCGTCTTGT ATCAATATAA CTCCAGAAAC TCCTACTGTC
    b-ugd21 TCAAATCGTA TATGTTTTTG GGCGTCTTGT ATCAATATAA CTCCAGAAAC TCCTACTGTC
    b-ugd10 TCAAATCGTA TATGTTTTTG GGCGTCTTGT ATCAATATAA CTCCAGAAAC TCCTACTGTC
   b-ukma24 TCAAATCGTA TATGTTTTTG GGCGTCTTGT ATCAATATAA CTCCAGAAAC TCCTACTGTC
    b-ugd29 TCAAATCGTA TATGTTTTTG GGCGTCTTGT ATCAATGTAA CTCCAGAAAC TCCTACTGTC
     b-ug52 TCAAATCGTA TATGTTTTTG GGCGTCTTGT ATCAATATAA CTCCAGAAAC TCCTACTGTC
     b-g413 TCAAATCGTA TATGTTTTTG GGCGTCTTGT ATCAATATAA CTCCAGAAAC TCCTACTGTC
     b-g482 TCAAATCGTA TATGTTTTTG GGCGTCTTGT ATCAATATAA CTCCAGAAAC TCCTACTGTC
    b-ugd13 TCAAATCGTA TATGTTTTTG GGCGTCTTGT ATTAATATAA CTCCAGAAAC TCCTACTGTC
     b-ugd3 TCAAATCGTA TATGTTTTTG GGCGTCTTGT ATTAATATAA CTCCAGAAAC TCCTACTGTC
    b-ugd15 TCAAATCGTA TATGTTTTTG GGCGTCTTGT ATTAATATAA CTCCAGAAAC TCCTACTGTC
     b-g71  TCAAATCGTA TATGTTTTTG GGCGTCTTGT ATCAATATAA CTCCAGAAAC TCCTACTGTC
     b-g91  TCAAATCGTA TATGTTTTTG GGCGTCTTGT ATCAATCTAA CTCCTGAAAC TCCTACTCTC
     c-gk17 TCAAATCGTC TCTGTTTTTG GGCGCGTTGT GCCAATATAA CTCCAGAAAC TCATACTGTA
     c-iap3 TCAAATCGTC TATGTTTTTG GGCGCGTTGT GCCAATATAA CTCCAGAAAC TCATACTGTA
     c-erla TCAAATCGTC TATGTTTTTC GGCGCGTTGT GCCAATCTAA CTCCAGAAAC TCATACTGTA
    c-icam1 TCAAATCGTC TATGTTTTTC GGCGCGTTGT GCCAATCTAA CTCCAGAAAC TCATACTGTA
    c-ukma1 TCAAATCGTC TATGTTTTTC GGCGCGTTGT GCCAATATAA CTCCAGAAAC TCATACTGTA
     c-iap2 TCAAATCGTC TATGTTTTTC GGCGCGTTGT GCCAATAGAA CTCCAGAAAC TCATACTGTA
     c-bbg1 TCAAATCGTC TATGTTTTTG GGCGCGTTGT GCCAATATAA CTCCAGAAAC TCATACTGTA
    c-ukma8 TCAAATCGTC TATGTTTTTG GGCGCGTTGT GCCAATATAA CTCCAGAAAC TCATACTGTA
      c-bc2 TCAAATCCTC TATGTTTTTC GGCGCGTTGT GCCAATATAA CTCTAGAAAC TGATACTGTA
    c-ukma3 TCAAATCGTC TATGTTTTTC GGCGCGTTGT GCCAATATAA CTCTAGAAAC TAATACTGTA
     c-ive1 TCAAATCGTC TATGTTTTTC GGCGCGTTGT GCCAATATAA CTCTAGAAAC TCATACTGTA
    c-ukb22 TCAAATCGTC TATGTTTTTC GGCGCGTTGT GCCAATATAA CTCTAGAAAC TCATACTGTA
     c-gk18 TCAAATCGTC TATGTTTTTC GGCGCGTTGT GCCAATATAA CTCCAGAAAC TCATACTGTA
  c-k1-8-dem TCAAATCGTA TATGTTTTTG GGCGCGTTGT GCCAATATAA CTCCAGAAAC TCATACTGTA
    c-ugd23 TCAAATCGTA TATGTTTTTG GGCGCGTTGT GCCAATATAA CTCCAGAAAC TCATACTGTA
    a-ugd16 TCAAATCGTA TATGTTTTTG GGCGCGTTGT GCCAATATAA CTCCAGAAAC TCATACTGTA
    a-ugd18 TCAAATCGTA TATGTTTTTG GGCGCGTTGT GCCAATATAA CTCCAGAAAC TCATACTGTA
    a-ugd24 TCAAATCGTA TATGTTTTTG GGCGCGTTGT GCCAATATAA CTCCAGAAAC TCATACTGTA
    a-ug374 TCAAATCGTA TATGTTTTTG GGCGCGTTGT GCCAATATAA CTCCAGAAAC TCATACTGTA
     a-ugd4 TCAAATCGTA TATGTTTTTG GGCGCGTTGT GCCAATATAA CTCCAGAAAC TCATACTGTA
    a-bcbl1 TCAAATCGTA TATGTTTTTG GGCGCGTTGT GCCAATATAA CTCCAGAAAC TCATACTGTA
    a-uka13 TCAAATCGTA TATGTTTTTG GGCTCGTTGT GCCAATCTAA CTCCAGAAAC TCATACTGTA
    a-uka21 TCAAATCGTA TATGTTTTTG GGCTCGTTGT GCCAATATAA CTCCAGAAAC TCATACTGTA
    a-ukma4 TCAAATCGTA TATGTTTTTG GGCTCGTTGT GCCAATATAA CTCCAGAAAC TCATACTGTA
      a-bc1 TCAAATCGTA TATGTTTTTG GGCGCGTTGT GCCAATATAA CTCCAGAAAC TCATACTGTA
     a-ema7 TCAAATCGTA TATGTTTTTG GGCGCGTTGT GACAATATAA CTCCAGAAAC TCATACTGTA
    a-bcblr TCAAATCGTA TATGTTTTTG GGCGCGTTGT GCCAATATAA CTCCAGAAAC TCATACTGTA
    a-ukb12 TCAAATCGTA TATGTTTTTG GGCGCGTTGT GCCAATATAA CTCCAGAAAC TCATACTGTA
    a-ug111 TCAAATCGTA TATGTTTTTG GGCGCGTTGT GCCAATATAA CTCCAGAAAC TCATACTGTA
     a-iap1 TCAAATCGTA TATGTTTTTG GGCGCGTTGT GCCAATATAA CTCCAGAAAC TCATACTGTA
    a-ukc12 TCAAATCGTA TATGTTTTTG GGCGCGTTGT GCCAATATAA CTCCAGAAAC TCAGACTGTA
c-k1-43-ber TCAAATCCTA TATGTTTTTG GCCGCGTTGT GCCAATATAA CTCCAGAAAC TCATACTGTA
    a-ugd12 TCAAATCGTA TATGTTTTTG GGCGCGTTGT GCCAATATAA CTCCAGAAAC TCATACTGTA
  Consensus TCAAATC-T- T-TGTTTTT- G-C---TTGT ---AAT--A- CTC--GAAAC T---ACT-T-
```

```
              541                                                       600
   d1-tks10   TCTGTCAGGA GTACTACAGG CTTTAAAACT CATACTGTAT CTGTCAGGAG TACTACAGGC
    d2-zks3   TCTGTCAGCA GTACTACAGG CTTTAA.... .......... .......... ..........
     b-ugd1   TCCGCCAGCA GTACTACGGC CTTTAA.... .......... .......... ..........
    b-ugd26   TCCGCCAGCA GTACTACGGC CTTTAA.... .......... .......... ..........
    b-ugd19   TCCGCCAGCA GTACTACGGC CTTTAA.... .......... .......... ..........
     b-ugd7   TCCGCCAGCA GTACTACGGC CTTTAA.... .......... .......... ..........
     b-ugd2   TCCGCCAGCA GTACTACGGC CATTAA.... .......... .......... ..........
    b-ugd21   TCCGCCAGCA GTACTACGGC CATTAA.... .......... .......... ..........
    b-ugd10   TCCGCCAGCA GTACTATGGC CGTTAA.... .......... .......... ..........
   b-ukma24   TCCGCCAGCA GTACTATGGC CGTTAA.... .......... .......... ..........
    b-ugd29   TCCGCCAGCA GTCCTATGGC CGTTAA.... .......... .......... ..........
     b-ug52   TCCGCCAGCA GTACTACGGC CTTTAA.... .......... .......... ..........
     b-g413   TCCTCCAGCA GTACTACGGC CTTTGA.... .......... .......... ..........
     b-g482   TCCTCCAGCA GTACTACGGC CTTTAA.... .......... .......... ..........
    b-ugd13   TCCGCCAGCA GTACTACGGC CTTTAA.... .......... .......... ..........
     b-ugd3   TCCGCCAGCA GTACTACGGC CTTTAA.... .......... .......... ..........
    b-ugd15   TCCGCCAGCA GTACTACGGC CTTTAA.... .......... .......... ..........
      b-g71   TCCGCCAGCA GTACTACGGC CTTTAA.... .......... .......... ..........
      b-g91   TCCGCCAGCA GTACTACGGC CTTTAA.... .......... .......... ..........
     c-gk17   TCTGTCAGCA GTACTACAGG CTTTAG.... .......... .......... ..........
     c-iap3   TCTGTCAGCA GTACTACAGG CTTTAG.... .......... .......... ..........
     c-erla   TCTGTCAGCA GTACTACAGG CTTTAG.... .......... .......... ..........
    c-icam1   TCTGTCAGCA GTACTACAGG CTTTAC.... .......... .......... ..........
    c-ukma1   TCTGTCAGCA GTACTACAGG CTTTAG.... .......... .......... ..........
     c-iap2   TCTGTCAGCA GTACTACAGG CTTTAG.... .......... .......... ..........
     c-bbg1   TCTGTCAGCA GTACTACAGG CTTTAG.... .......... .......... ..........
    c-ukma8   TCTGTCAGCA GTACTACAGG CTTTAG.... .......... .......... ..........
      c-bc2   TCTGTCAGCA GTACTACAGG CTTTAG.... .......... .......... ..........
    c-ukma3   TCTGTCAGCA GTACTACAGG CTTTAG.... .......... .......... ..........
     c-ive1   TCTGTCAGCA GTACTACAGG CTTTAG.... .......... .......... ..........
    c-ukb22   TCTGTCAGCA GTACTACAGG CTTTAG.... .......... .......... ..........
     c-gk18   TCTGTCAGCA GTACTACAGG CTTTAG.... .......... .......... ..........
  c-k1-8-dem  TCTGTCAGCA GTACTACAGG CTTTGG.... .......... .......... ..........
    c-ugd23   TCTGTCAGCA GTACTACAGG CTTTGG.... .......... .......... ..........
    a-ugd16   TCTGTCAGCA GTACTACAGG CTTTAG.... .......... .......... ..........
    a-ugd18   TCTGTCAGCA GTACTACAGG CTTTAG.... .......... .......... ..........
    a-ugd24   TCTGTCAGCA GTACTACAGG CTTTAG.... .......... .......... ..........
    a-ug374   TCTGTCAGCA GTACTACAGG CTTTAG.... .......... .......... ..........
     a-ugd4   TCTGTCAGCA GTACTACAGG CTTTAG.... .......... .......... ..........
    a-bcbl1   TCTGTCAGCA GTACTACAGG CTTTAG.... .......... .......... ..........
    a-uka13   TCTGTCAGCA GTACTACAGG CTTTAG.... .......... .......... ..........
    a-uka21   TCTGTCAGCA GTACTACAGG CTTTAG.... .......... .......... ..........
    a-ukma4   TCTGTCAGCA GTACTACAGG CTTTAG.... .......... .......... ..........
      a-bc1   TCTGTCAGCA GTACTACAGG CTTTAG.... .......... .......... ..........
     a-ema7   TCTGTCAGCA GTACTACAGG CTTTAG.... .......... .......... ..........
    a-bcblr   TCTGTCAGCA GTACTACAGG CTTTAG.... .......... .......... ..........
    a-ukb12   TCTGTCAGCA GTACTACAGG CTTTAG.... .......... .......... ..........
    a-ug111   TCTGTCAGCA GTACTACAGG CTTTAG.... .......... .......... ..........
     a-iap1   TCTGTTAGCA GTACTACAGG CTTTAG.... .......... .......... ..........
    a-ukc12   TCTGTCAGCA GTACTACAGG CTTTAG.... .......... .......... ..........
c-k1-43-ber   TCTGTCAGTA GTACTACAGG CTTTAG.... .......... .......... ..........
    a-ugd12   TCTGTCAGCA GTACTA.... .......... .......... .......... ..........
  Consensus   TC----AG-A GT-CTA---- ---------- ---------- ---------- ----------
```

```
                601                                                        660
   d1-tks10  TTTACAACAT TCAGTACTAA TAGATTAGTG AACATAATCC CTGCAACCAC ACATGCTGTA
    d2-zks3  .....AACAT TCAGTACTAA TAGATTAGTG AACATAATCC CTGCAACCAC ACATGCTGTA
     b-ugd1  .....AATGT CAAGAACTAA TGGATTACTG AAAATAATCC CTGCAACCAC ACATGCTGCA
    b-ugd26  .....AATGT CAAGAACTAA TGGATTACTG AAAATAATCC CTGCAACCAC ACATGCTGCA
    b-ugd19  .....AATGT CAAGAACTAA TGGATTACTG AAAATAATCC CTGCAACCAC ACATGCTGCA
     b-ugd7  .....AATGT CAAGAACTAA TGGATTACTG AAAATAATCC CTGCAACCAC ACATGCTGCA
     b-ugd2  .....AATGT CAAGAACTAA TGGATTACTG AAAATAATCC CTGCAACCAC ACATGCTGCA
    b-ugd21  .....AATGT CAAGAACTAA TGGATTACTG AAAATAATCC CTGCAACCAC ACATGCTGCA
    b-ugd10  .....ACTAT TAAGAACTAA TGGATTACTG AAAATAATCC CTGCAACCAC ACATGCTGCA
   b-ukma24  .....AGTAT TAAGAACTAA TGGATTACTG AAAATAATCC CTGCAACCAC ACATGCTGCA
    b-ugd29  .....AGTAT TAAGAACTAA TGGATTACTG AAAATAATCC CTGCAACCAC ACATGCTGCA
     b-ug52  .....AATGT CAAGAACTAA TGGATTACTG AAAATAATCC CTGCAACCAC ACATGCTGCA
     b-g413  .....AACAT TAAGAACTAA TGGATTACTG AAAATAATCC CTGCAACCAC ACATGCTGCA
     b-g482  .....AACAT TAAGAACTAA TGGATTACTG AAAATAATCC CTGCAACCAC ACATGCTGCA
    b-ugd13  .....AACAT TAACAACTAA TGGATTACTG AAAATAATCC CTGCAACCAC ACATGCTGCA
     b-ugd3  .....AACAT TAACAACTAA TGGATTACTG AAAATAATCC CTGCAACCAC ACATGCTGCA
    b-ugd15  .....AACAT TAACAACTAA TGGATTACTG AAAATAATCC CTGCAACCAC ACATGCTGCA
     b-g71   .....AAAAT TAAGAAATAA TGGATTACTG AAAATAATCC CTGCAACCAC ACATGCTGCA
     b-g91   .....AACAT TAAGAACTAA TGGATTACTG AAAATAATCC CTGCAACCAC ACATGCTGCA
     c-gk17  .....AACAT TCAGTACTCA TAG.......  ........CC ATACAACCAC ACATGCTGTA
     c-iap3  .....AACAT TCAGTACTCA TAG.......  ........CC ATACAACCAC ACATGCTGTA
     c-erla  .....AACAT TCAGTACTAA TAG.......  ........AC GTGGAACCAC ACTTGATGTA
    c-icam1  .....AACAT TCAGTACTAA TAG.......  ........AC GTGGAACCAC ACTTGATGTA
    c-ukma1  .....AACAT TCAGTACTAA TAG.......  ........AC ATGCAGCCAC ACATGATATA
     c-iap2  .....AACAT TCAGTACTAA TAG.......  ........CC ATACAACCGC ACATAATGTA
     c-bbg1  .....AACAT TCAGTACTAA TAG.......  ........CC ATGCAACCAA ACATGATGTA
    c-ukma8  .....AACAT TCAGTACTAA TAG.......  ........CC ATGCAACCAA ACATGATGTA
      c-bc2  .....AACAT TCAGTACTAA TAG.......  ........CG CTGCAACCAC ACATGATGTA
    c-ukma3  .....AACAT TCAGTACTAA TAG.......  ........CG ATGCAACCAC ACATGATATA
     c-ive1  .....AACAT TCAGTACTAA TAG.......  ........CC ATGCAACCAC ACATGATGAA
    c-ukb22  .....AACAT TCAGTACTAA TAG.......  ........CC ATGCAACCAC ACATGATGTA
     c-gk18  .....AACAT T.........  ..........  .......... .TGCAACCGC ACCTACTCTA
  c-k1-8-dem .....AACAT TCAGTACTCA TAGATTAGTG AACAGAATCC ATGCAACCAC ACATGATGTA
    c-ugd23  .....AACAT TGAGTACTCA TAGCTTAGTG AACAGAATCC ATGCAACCAC ACATGATGTA
    a-ugd16  .....AACAG TCAGTACTAA TAGCTTAGTG AACATAATCC ATGCAACCAA CCATGATGTA
    a-ugd18  .....AACAG TCAGTACTAA TAGCTTAGTG AACATAATCC ATGCAACCAA CCATGATGTA
    a-ugd24  .....AACAG TCAGTACTAA TAGCTTAGTG AACATAATCC ATGCAACCAC ACATGATGTA
    a-ug374  .....AACAT TCAGTACTAA TAGCTTAGTG AACATAATCC ATGCAACCAC ACATGATGTA
     a-ugd4  .....AAATT TAAGTACTCA TAGCTTAGTG AACATAATCC ATGCAACCAC ACATCATGTA
    a-bcbl1  .....AACAT TCAGTACTAA TAGCTTAGTG AACATAATCC ATGCAACCAC ACATAAAGTA
    a-uka13  .....AACAT TCAGTACTAA TAGGTTAGTG AACATAATCC ATGCAACCAC ACATGATGTA
    a-uka21  .....AACAT TCAGTACTAA TAGGTTAGTG AACATAATCC ATGCAACCAC ACATGATGTA
    a-ukma4  .....AACAT TCAGTACTAA TAGGTTAGTG AACATAATCC ATGCAACCAC ACATGATGTA
      a-bc1  .....AACAT TGAGTACTAA TAGCTTAGTG AAGATAATCC ATGCAACCAC ACGTGATGTA
     a-ema7  .....AACAT TGAGTACTAA TAGCTTAGTG AACATAATCC ATGCAACCAC ACATGATGTA
    a-bcblr  .....AACAT TCAGTACTAA TAGCTTAGTG AACATAATCC ATGCAACCAC ACATGATGTA
    a-ukb12  .....AACAT TCAGTACTAA TAGCTTAGTG AACATAATCC ATGCAACCAC ACATGATGTA
    a-ug111  .....AACAT TCAGTACTAA TAGCTTAGTG AACATAATCC ATGCAACCAC ACATAATGTA
     a-iap1  .....AACAT TCAGTACTAA TAGCTTAGTG AACATAATCC ATGCAACCAC ACATGATGCA
    a-ukc12  .....AACAT TCAGTACTAA TAGCTTAGTG AACATAATCC ATG.......  .....ATGTA
c-k1-43-ber  .....AACAT TCAGTACTAA TAGCATAGTG AACATAACCC ATGCAACCAC ACGTGCTGTA
    a-ugd12  ..........  ..........  ..........  ....TAATCC ATGCAACCAC ACATGATGTA
   Consensus ----------  ----------  ----------  ---------- -T-------- ---------A
```

```
              661                                                                720
    d1-tks10  GTTGTAGTGG AAAAAGTAAA ATCTCTACAT CCACATATTG AAGTGCCTTT TCTTGTATTT
     d2-zks3  GTTGTAGTGG AAGAACTAAA ATCTAGAAAT CCATATATTA AAGTGCATTT TCTTATATTA
      b-ugd1  GTTGCAGTGG AAGAAGTAAA ATCTACAAAT CCACATATTC AAGTGCCTTT TCTTGTATTT
     b-ugd26  GTTGCAGTGG AAGAAGTAAA ATCTACAAAT CCACATATTC AAGTGCCTTT TCTTGTATTT
     b-ugd19  GTTGCAGTGG AAGAAGTAAA ATCTACAAAT CCACATATTC AAGTGCCTTT TCTTGTATTT
      b-ugd7  GTTGCAGTGG AAGAAGTAAA ATCTACAAAT CCACATATTC AAGTGCCTTT TCTTGTATTT
      b-ugd2  GTTGCAGTGG AAGAAGTAAA ATCTACAAAT CCACATATTC AAGTGCCTTT TCTTGTATTT
     b-ugd21  GTTGCAGTGG AAGAAGTAAA ATCTACAAAT CCACATATTC AAGTGCCTTT TCTTGTATTT
     b-ugd10  GTTGCAGTGG AAGAAGTAAA ATCTACAAAT ACACATATTC AAGTGCCTTT TCTTGTATTT
    b-ukma24  GTTGCAGTGG AAGAAGTAAA ATCTACAAAT ACACATATTC AAGTGCCTTT TCTTGTATTT
     b-ugd29  GTTGCAGTGG AAGAAGTAAA ATCTACAAAT ACACATATTC AAGTGCCTTT TCTTGTATTT
      b-ug52  GTTGCAGTGG AAGAAGTAAA ATCTACAAAT CCACATATTC AAGTGCCTTT TCTTGTATTT
      b-g413  GTTGCAGTGG AAGAAGTAAA ATCTACAAAT CCACATATTC AAGTGCCTTT TCTTGTATTT
      b-g482  GTTGCAGTGG AAGAAGTAAA ATCTACAAAT CCACATATTC AAGTGCCTTT TCTTGTATTT
     b-ugd13  GTTGCAGTGG AAGAAGTAAA ATCTACAAAT CCACATATTC AAGTGCCTTT TCTTGTATTT
      b-ugd3  GTTGCAGTGG AAGAAGTAAA ATCTACAAAT CCACATATTC AAGTGCCTTT TCTTGTATTT
     b-ugd15  GTTGCAGTGG AAGAAGTAAA ATCTACAAAT CCACATATTC AAGTGCCTTT TCTTGTATTT
      b-g71   GTTGCAGTGG AAGAAGTAAA ATCTACAAAT CCACATATTC AAGTGCCTTT TCTTGTATTT
      b-g91   GTTGCAGTGG AAGAAGTAAA ATCTACAAAT CCACATATTC AAGTGCCTTT TCTTGTATTT
      c-gk17  GTTGTAGTGA AAGAAGCAAA ATTTACAAAT CCACATATTG AAGTGCCTTT TCTTGTATTT
      c-iap3  GTTGTAGTGA AAGAAGCAAA ATTTACAAAT CCACATATTG AAGTGCCTTT TCTTGTATTT
      c-erla  CTTGTAATGA AAGAAGCAAA ATCTACAAAT CTACATATTC AAGTGCATTT TCTTGTATTT
     c-icam1  CTTGTAATGA AAGAAGCAAA ATCTACAAAT CTACATATTC AAGTGCATTT TCTTGTATTT
     c-ukma1  CTTGTAATGA AAGAAGCAAA ATCTACAAAT CTACATATTC AAGTGCATTT TCTTGTATTT
      c-iap2  CTTGTAATGA AAGAAGCAAA ATCTACAAAT CTACATATTC AAGTGCATTC TCTTGTATTT
      c-bbg1  GTTGTAGTAA AAGAAGCAAA ATTTACAAAT CCACATATTG AAGTGCCTTT TCTTGTATTT
     c-ukma8  GTTGTAGTGA AAGAAGCAAA ATTTACAAAT CCACATATTG AAGTGCCTTT TCTTGTATTT
       c-bc2  CTTGTAATGA AAGAAGCCAA ATCTACAAAT CTACATATTC AAGTGCATTT TATTGTATTT
     c-ukma3  CTTGTAATGA AAGAAGCCAA ATCTACAAAT CTACATATTC AAGTGCATTT TCTTGTATTT
      c-ive1  CTTGTAATGA AAGAAGCCAA ATCTACAAAT CTACATATTC AAGTGCATTT TCTTGTATTT
     c-ukb22  CTTGTAATGA AAGAAGCCAA ATCTACAAAT CTACCTATTC AAGTGCATTT TTTTGTATTT
      c-gk18  TTTGTAATGA AAGAAGTAAA ATCTACATAT CTATATATTC AAGAGCATTT GCTTGTATTT
   c-k1-8-dem GTTGTAGTGA AAGAAGCAAA ATCTACAAAT CCACATATTG AAGTGCCTTT TCTTGTATTT
     c-ugd23  GTTGTAGTGA AAGAAGCAAA ATCTACAAAT CCACATATTG AAGTGCCTTT TCTTGTATTT
     a-ugd16  GTTGTAGTGA AAGAAGCAAA ATCTACAAAT CCACATATTG AAGTGCCTTT TCTTGTATTT
     a-ugd18  GTTGTAGTGA AAGAAGCAAA ATCTACAAAT CCACATATTG AAGTGCCTTT TCTTGTATTT
     a-ugd24  GTTGTAGTGA AAGAAGCAAA ATCTACAAAT CCACATATTG AAGTGCCTTT TCTTGTATTT
     a-ug374  GTTGTAGTGA AAGAAGCAAA ATCTACAAAT CCACATATTG AAGTGCCTTT TCTTGTATTT
      a-ugd4  GTTGTAGTGA AAGAAGCAAA ATCTACAAAT CCACATATTG AAGTGCCTTT TCTTGTATTT
     a-bcbl1  GTTGTAGTGA AAGAAGCAAA ATCTACAAAT TCACATATTG AAGTGCATTT TCTTGTATTT
     a-uka13  GTTGTACTGA AAGAAGCAAA ATCTACACGT TTTCATATTG AACTGCATTT TCTTGTATTT
     a-uka21  GTTGTACTGA AAGAAGCAAA ATCTACACGT TTTCATATTG AACTGCATTT TCTTGTATTT
     a-ukma4  GTTGTACTGA AAGAAGCAAA ATCTATACGT TTTCATATTG AACTGCATTT TCTTGTATTT
       a-bc1  GTTGTAGTGA AAGAAGCAAA ATCTACACAT TTTCATATTG AAGTGCATTT TCTTGTATTT
      a-ema7  GTTGTAGTGA AAGAAGCAAA ATCTACACAT TTTCATATTG AACTGCATTT TCTTGTATTT
     a-bcblr  GTTGTAGTGA AAGAAGCAAA ATCTACACAT TTTCATATTG AACTGCATTT TCTTGTATTT
     a-ukb12  GTTGCAGTGA AAGAAGCAAA ATCTACACAT TTTCATATTG AACTGCATTT TCTTGTATTT
     a-ug111  GTTGTAGTGA AAGAAGCAAA ATCTACACAT TTTCATATTG AACTGCATTT TCTTGTATTT
      a-iap1  GTTGTAGTGA AAGAAGCAAA ATCTACACGT TTTCATATTG AAGTGCATTT TCTTGTATTT
     a-ukc12  GTTGTAGTGA AAGAAGCAAA ATCTACACAT TTTCATATTG AACTGCATTT TCTTGTATTT
   c-k1-43-ber GTTGTAGTGA AAGAAGCAAA ATCTACAAAT CCACATATTG AAGTGCCTTT TCTTGTATTT
     a-ugd12  GTTGTAGTGA AAGAAGCAAA ATCTACAAAT CCACATATTG AAGTGCCTTT TCTTGTATTT
   Consensus  -TTG-A-T-- AA-AA---AA AT-T--A--T -----TATT- AA--GC-TT- --TT-TATT-
```

```
                721                                                                 780
   d1-tks10  ATGACACTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTAT TATCTTTGCC
    d2-zks3  ATGACGCTGG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTAT TATCTTTGCC
     b-ugd1  ATGACGCTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTCT TATCTTTACC
    b-ugd26  ATGACGCTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTCT TATCTTTACC
    b-ugd19  ATGACGCTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTCT TATCTTTACC
     b-ugd7  ATGACGCTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTCT TATCTTTACC
     b-ugd2  ATGACGCTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTCT TATCTTTACC
    b-ugd21  ATGACGCTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTCT TATCTTTACC
    b-ugd10  ATGACGCTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTCT TATCTTTACC
   b-ukma24  ATGACGCTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTCT TATCTTTACC
    b-ugd29  ATGACGCTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTCT TATCTTTACC
     b-ug52  ATGACGCTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTCT TATCTTTACC
     b-g413  ATGACGCTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTGT TATCTTTACC
     b-g482  ATCACGCTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAGCTGT TATCTTTACC
    b-ugd13  ATGACGCTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTGT TATCTTTACC
     b-ugd3  ATGACGCTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTGT TATCTTTACC
    b-ugd15  ATGACGCTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTGT TATCTTTACC
     b-g71   ATGACGCTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTGT TATCTTTACC
     b-g91   ATGACGCTCG TAGCTCTCAT AGGAACCATG TGTGGTATCT TCGGAACTCT TATCTTTGCC
     c-gk17  ATGACACTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTAT TATCTTTGCC
     c-iap3  ATGACACTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTAT TATCTTTGCC
     c-erla  ATGACACTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTAT TATCTTTGCC
    c-icam1  ATGACACTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTAT TATCTTTGCC
    c-ukma1  ATGACACTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTAT TATCTTTGCC
     c-iap2  ATGACACTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTAT TATCTTTGCC
     c-bbg1  ATGACACTCG TAGCTCTAAT AGGAACCATG TGTGGTATCT TAGGAACTAT TATCTTTGCC
    c-ukma8  ATGACACTCG TAGCTCTAAT AGGAACCATG TGTGGTATCT TAGGAACTAT TATCTTTGCC
      c-bc2  ATGACACTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTAT TATCTTTGCC
    c-ukma3  ATGACACTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTAT TATCTTTGCC
     c-ive1  ATGACACTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTAT TATCTTTGCC
    c-ukb22  ATGACACTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTAT TATCTTTGCC
     c-gk18  ATGACACTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTAT TATCTTTGCC
  c-k1-8-dem ATGACACTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTAT TATCTTTGCC
    c-ugd23  ATGACACTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTAT TATCTTTGCC
    a-ugd16  ATGACACTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTAT TATCTTTGCC
    a-ugd18  ATGACACTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTAT TATCTTTGCC
    a-ugd24  ATGACACTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTAT TATCTTTGCC
    a-ug374  ATGACACTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTAT TATCTTTGCC
     a-ugd4  ATGACACTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTAT TATCTTTGCC
    a-bcbl1  ATGACACTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTAT TATCTTTGCC
    a-uka13  ATGACACTCG TAGCTCTGAT AGGAACCATG TGTGGTATCG TAGGAACTAT TATCTTTTTC
    a-uka21  ATGACACTCG TAGCTCTGAT AGGAACCATG TGTGGTATCG TAGGAACTAT TATCTTTTTC
    a-ukma4  ATGACACTCG TAGCTCTGAT AGGAACCATG TGTGGTATCG TAGGAACTAT TATCTTTTTC
      a-bc1  ATGACACTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTAT TATCTTTGCC
     a-ema7  ATGACACTCG TAGCTCTGAT AGGAACCGTG TGTGGTATCT TAGGAACTAT TATCTTTGCC
    a-bcblr  ATGACACTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTAT TATCTTTGCC
    a-ukb12  ATGACACTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TGGGAACTAT TATCTTTGCC
    a-ug111  ATGACACTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTAT TATCTTTGCC
     a-iap1  ATGACACTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTAT TATCTTTGCC
    a-ukc12  ATGACACTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTAT TATCTTTGCC
  c-k1-43-ber ATGACACTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTAT TATCTTTGCC
    a-ugd12  ATGACACTCG TAGCTCTGAT AGGAACCATG TGTGGTATCT TAGGAACTAT TATCTTTGCC
  Consensus  AT-AC-CT-G TAGCTCT-AT AGGAACC-TG TGTGGTATC- T-GGA-CT-T TATCTTT--C
```

```
              781                                                       840
   d1-tks10   CATTGTCAAA AACAAAGTGA CTCAAACAAA ACAGTGCAAC AACAATTGCG GGATTATTAT
    d2-zks3   CGTTGTCAAA AACAAAGTGA CTCAAACAAA ACAGTGCCAC AACAATTGCG GGATTATTAT
     b-ugd1   CATTGTCAAA AAAAAAGTGA CTCAAGCAAA ACAGGGCAAC AACAATTGCG GGATTATTAT
    b-ugd26   CATTGTCAAA AAAAAAGTGA CTCAAGCAAA ACAGGGCAAC AACAATTGCG GGATTATTAT
    b-ugd19   CATTGTCAAA AAAAAAGTGA CTCAAGCAAA ACAGGGCAAC AACAATTGCG GGATTATTAT
     b-ugd7   CATTGTCAAA AAAAAAGTGA CTCAAGCAAA ACAGGGCAAC AACAATTGCG GGATTATTAT
     b-ugd2   CATTGTCAAA AAAAAAGTGA CTCAAGCAAA ACAGGGCAAC AACAATTGCG GGATTATTAT
    b-ugd21   CATTGTCAAA AAAAAAGTGA CTCAAGCAAA ACAGGGCAAC AACAATTGCG GGATTATTAT
    b-ugd10   CATTGTCAAA AAAAAAGTGA CTCAAGCAAA ACAGGGCAAC AACAATTGCG GGATTATTAT
   b-ukma24   CATTGTCAAA AAAAAAGTGA CTCAAGCAAA ACAGGGCAAC AACAATTGCG GGATTATTAT
    b-ugd29   CATTGTCAAA AAAAAAGTGA CTCAAGCAAA ACAGGGCAAC AACAATTGCG GGATTATTAT
     b-ug52   CATTGTCAAA AAAAAAGTGA CTCAAGCAAA ACAGGGCAAC AACAATTGCG GGATTATTAT
     b-g413   CATTGTCAAA AAAAAAGTGA CTCAAGCAAA ACAGGGCAAC AACAATTGCG GGATTATTAT
     b-g482   CATTGTCAAA AAAAAAGTGA CTCAAGCAAA ACAGGGCAAC AACAATTGCG GGATTATTAT
    b-ugd13   CATTGTCAAA AAAAAAGTGA CTCAAGCAAA ACAGGGCAAC AACAATTGCG GGGTTATTAT
     b-ugd3   CATTGTCAAA AAAAAAGTGA CTCAAGCAAA ACAGGGCAAC AACAATTGCG GGGTTATTAT
    b-ugd15   CATTGTCAAA AAAAAAGTGA CTCAAGCAAA ACAGGGCAAC AACAATTGCG GGATTATTAT
     b-g71    CATTGTCAAA AAAAAAGTGA CTCAAGCAAA ACAGGGCAAC AACAATTGCC GGATTATTAT
     b-g91    CATTGTCAAA AACAAAGTGA CTCAAGCAAC ACAGGGCAAC AACAATTGCG GGATTATTAT
     c-gk17   CATTGTCAAA AACAAAGTGA CTCAAACAAA ACAGTGCCAC AACAATTGCG GGATTATTAT
     c-iap3   CATTGTCAAA AACAAAGTGA CTCAAACAAA ACAGTGCCAC AACAATTGCG GGATTATTAT
     c-erla   CATTGTCAAA AACAACGTGA CTCAAACAAA ACAGTGCCAC AACAATTGCA GGATTATTAT
     c-icam1  CATTGTCAAA AACAACGTGA CTCAAACAAA ACAGTGCCAC AACAATTGCA GGATTATTAT
     c-ukma1  CATTGTCAAA AACAACGTGA CTCAAACAAA ACAGTGCCAC AACAATTGCA GGATTATTAT
     c-iap2   CATTGTCAAA AACAACGTGA CTCAAACAAA ACAGTGCCAC AACAATTGCA GGATTATTAT
     c-bbg1   CATTGTCAAA AACAAAGTGA CTCAAACAAA ACAGTGCCAC AACAATTGCG GGATTATTAT
     c-ukma8  CATTGTCAAA AACAAAGTGA CTCAAACAAA ACAGTGCCAC AACAATTGCG GGATTATTAT
      c-bc2   CATTGTCAAA AACAACGTGA CTCAAACAAA ACAGTGCCAC AACAATTGCA GGATTATTAT
     c-ukma3  CATTGTCAAA AACAAAGTGA CTCAAACAAA ACAGTGCCAC AACAATTGCG GGATTATTAT
     c-ive1   CATTGTCAAA AACAAAGTGA CTCAAACAAA ACAGTGCCAC AACAATTGCG GGATTATTAT
     c-ukb22  CATTGTCAAA AACAAAGTGA CTCAAACAAA ACAGTGCCAC AACAATTGCG GGATTATTAT
     c-gk18   CATTGTCAAA AACAACGTGA CTCAAACAAA ACAGTGCCAC AACAATTGCA GGATTATTAT
   c-k1-8-dem CATTGTCAAA AACAAAGTGA CTCAAAAAAA ACAGTGCCAC AACAATTGCG GGATTATTAT
    c-ugd23   CATTGTCAAA AACAAAGTGA CTCAAACAAA ACAGTGCCAC AACAATTGCG GGATTATTAT
    a-ugd16   CATTGTCAAA AACAACGTGA CTCAAACAAA ACAGTGCCAC AACAATTGCA GGATTATTAT
    a-ugd18   CATTGTCAAA AACAACGTGA CTCAAACAAA ACAGTGCCAC AACAATTGCA GGATTATTAT
    a-ugd24   CATTGTCAAA AACAACGTGA CTCAAACAAA ACAGTGCCAC AACAATTGCA GGATTATTAT
    a-ug374   CATTGTCAAA AACAACGTGA CTCAAACAAA ACAGTGCCAC AACAATTGCA GGATTATTAT
     a-ugd4   CATTGTCAAA AACAACGTGA CTCAAACAAA ACAGTGCCAC AACAATTGCA GGATTATTAT
    a-bcbl1   CATTGTCAAA AACAACGTGA CTCAAACAAA ACAGTGCCAC AACAATTGCA GGATTATTAT
    a-uka13   CATTGTCAAA AACAACGTGA CTCAAACAAA ACAGTGCCAC AACAATTGCA GGATTATTAT
    a-uka21   CATTGTCAAA AACAACGTGA CTCAAACAAA ACAGTGCCAC AACAATTGCA GGATTATTAT
    a-ukma4   CATTGTCAAA AACAACGTGA CTCAAACAAA ACAGTGCCAC AACAATTGCA GGATTATTAT
      a-bc1   CATTGTCAAA AACAACGTGA CTCAAACAAA ACAGTGCCAC AACAATTGCA GGATTATTAT
     a-ema7   CATTGTCAAA AACAACGTGA CTCAAACAAA ACAGTGCCAC AACAATTGCA GGATTATTAT
    a-bcblr   CATTGTCAAA AACAACGTGA CTCAAACAAA ACAGTGCCAC AACAATTGCA GGATTATTAT
    a-ukb12   CATTGTCAAA AACAACGTGA CTCAAACAAA ACAGTGCCAC AACAATTGCA GGATTATTAT
    a-ug111   CATTGTCGAA AACAACGTGA CTCAAACAAA CCAGTGCCAC AACAATTGCA GGATTATTAT
     a-iap1   CATTGTCAAA AACAACGTGA CTCAAACAAA ACAGTGCCAC AACAATTGCA GGATTATTAT
    a-ukc12   CATTGTCAAA AACAACGTGA CTCAAACAAA ACAGTGCCAC AACAATTGCA GGATTATCAT
 c-k1-43-ber  CATTGTCAAA AACAAAGTGA CTCAAACAAA ACAGTGCCAC AACAATTGCG GGATTATTAT
    a-ugd12   CATTGTCAAA AACAACGTGA CTCAAACAAA ACAGTGCCAC AACAATTGCA GGATTATTAT
   Consensus  C-TTGTC-AA AA-AA-GTGA CTCAA--AA- -CAG-GC-AC AACAATTGC- GG-TTAT-AT
```

```
            841                                     879
   d1-tks10  TCCCTACACG ATTTTAACAC GGAAGACTAT ACGCAACCA
    d2-zks3  TCCCTACACG ATTTTATCAC GGAAGACTAT ATGCAACCA
     b-ugd1  TCCCTAGACT ATTTTCACAC GGAAGAGTAT ACGCAACCA
    b-ugd26  TCCCTAGACT ATTTTCACAC GGAAGAGTAT ACGCAACCA
    b-ugd19  TCCCTAGACT ATTTTCACAC GGAAGAGTAT ACGCAACCA
     b-ugd7  TCCCTAGACT ATTTTCACAC GGAAGAGTAT ACGCAACCA
     b-ugd2  TCCCTAGACT ATTTTCACAC GGAAGAGTAT ACGCAACCA
    b-ugd21  TCCCTAGACT ATTTTCACAC GGAAGAGTAT ACGCAACCA
    b-ugd10  TCCCTAGACT ATTTTCACAC GGAAGACTAT ACGCAACCA
   b-ukma24  TCCCTAGACT ATTTTCACAC GGAAGACTAT ACGCAACCA
    b-ugd29  TCCCTAGACT ATTTTCACAC GGAAGACTAT ACGCAACCA
     b-ug52  TCCCTAGACT ATTTTCACAC GGAAGACTAT ACGCAACCA
     b-g413  TCCCTAGACT ATTTTCACAC GGAAGACTAT ACGCAACCA
     b-g482  TCCCTAGACT ATTTTCACAC GGAAGACTAT ACGCAACCA
    b-ugd13  TCCCTAGACT ATTTTCACAC GGAAGAGTAT ACACAACCA
     b-ugd3  TCCCTAGACT ATTTTCACAC GGAAGAGTAT ACACAACCA
    b-ugd15  TCCCTAGACT ATTTTCACAC GGAAGAGTAT ACACAACCA
      b-g71  TCCCTAGACT ATTTTCACAC GGAAGAGTAT ACACAACCA
      b-g91  TCCCTAGACT ATTTTCACAC GGAAGAGTAT ACGCAACCA
     c-gk17  TCCCTACACG ATTTGTGCAC GGAAGACTAT ACGCAACCA
     c-iap3  TCCCTACACG ATTTGTGCAC GGAAGACTAT ACGCAACCA
     c-erla  TCCCTACACG ATTTGTGCAC GGAAGACTAT ACGCAACCA
    c-icam1  TCCCTACACG ATTTGTGCAC GGAAGACTAT ACGCAACCA
    c-ukma1  TCCCTACACG ATTTGTGCAC GGAAGACTAT ACGCAACCA
     c-iap2  TCCCTACACG ATTTGTGCAC GGAAGACTAT ACGCAACCA
     c-bbg1  TCCCTACACG ATTTGTGCAC GGAAGACTAT ACGCAACCA
    c-ukma8  TCCCTACACG ATTTGTGCAC GGAAGACTAT ACGCAACCA
      c-bc2  TCCCTACACG ATTTGTGCAC GGAAGACTAT ACGCAACCA
    c-ukma3  TCCCTACACG ATTTGTGCAC GGAAGACTAT ACGCAACCA
     c-ive1  TCCCTACACG ATTTGTGCAC GGAAGACTAT ACGCAACCA
    c-ukb22  TCCCTACACG ATTTGTGCAC GGAAGACTAT ACGCAACCA
     c-gk18  TCCCTACACG ATTTGTGCAC GGAAGACTAT ACGCAACCA
  c-k1-8-dem TCCCTACACG ATTTCTGCAC GGAAGACTAT ACGCAACCA
    c-ugd23  TCCCTACACG ATTTCTGCAC GGAAGACTAT ACGCAACCA
    a-ugd16  TCCCTACACG ATTTGTGCAC GGAAGACTAT ACGCAACCA
    a-ugd18  TCCCTACACG ATTTGTGCAC GGAAGACTAT ACGCAACCA
    a-ugd24  TCCCTACACG ATTTGTGCAC GGAAGACTAT ACGCAACCA
    a-ug374  TCCCTACACG ATTTGTGCAC GGAAGACTAT ACGCAACCA
     a-ugd4  TCCCTACACG ATTTGTGCAC GGAAGACTAT ACGCAACCA
    a-bcbl1  TCCCTACACG ATTTGTGCAC GGAAGACTAT ACGCAACCA
    a-uka13  TCCCTACACG ATTTGTGCAC GGAAGACTAT ACGCAACCA
    a-uka21  TCCCTACACG ATTTGTGCAC GGAAGACTAT ACGCAACCA
    a-ukma4  TCCCTACACG ATTTGTGCAC GGAAGACTAT ACGCAACCA
      a-bc1  TCCCTACACG ATTTGTGCAC GGAAGACTAT ACGCAACCA
     a-ema7  TCCCTACACG ATTTGTGCAC GGAAGACTAT ACGCAACCA
    a-bcblr  TCCCTACACG ATTTGTGCAC GGAAGACTAT ACGCAACCA
    a-ukb12  TCCCTACACG ATTTGTGCAC GGAAGACTAT ACGCAACCA
    a-ug111  TCCCTACACG ATTTGTGCAC GGAAGACTAT ACGCAACCA
     a-iap1  TCCCTACACG ATTTGTGCAC GGAAGACTAT ACGCAACCA
    a-ukc12  TCCCTACACG ATTTGTGCAC GGAAGACTAT ACGCAACCA
c-k1-43-ber TCCCTACACG ATTTGTGCAC GGAAGACTAT ACGCAACCA
    a-ugd12  TCCCTACACG ATTTGTGCAC GGAAGACTAT ACGCAACCA
  Consensus  TCCCTA-AC- ATTT---CAC GGAAGA-TAT A--CAACCA
```

B

```
                1                                                        60
     b-ugd2  LCSLVVWFPK LLSLHLPSFP HLCPGVHFTP YTLTCPSNRS LPISWYCNGT QLSRIRASTL
    b-ugd21  LCSLVVWFPQ LLSLHLPSFP HLCPGVHFTS YTLTCPSNRS LPISWYCNGT QLWRIRENTL
     b-ugd1  LCCLVVWFPK LLSLHLPSFP HLCPGVLFTP YTLTCPSNRS LPTSWYCNGT QLRRIRGSNL
    b-ugd26  LCSLVVWFPK LLSLHLPSFP HLCPGVLFTP YTLTCPSNRS LPISWYCNGT QLRRIRESTL
    b-ugd19  LCSLVVWFPK LLSLHLPSFP HLCPGVLFTP YTLTCPSNRS LPISWYCNGT RLWRIRESNL
     b-ugd7  LCSLVVCFPK LLSLHLPSFP HLCPGVLFTP YTLTCPSNRS LPISWYCNGT QLWRITDSTL
    b-ugd10  VCSLLVCFPK LLSLHLPSFP HLCPGVISTP YTLTCPSNRS LPISWYCNGT QLHRITASNL
   b-ukma24  LYGFLVCFPK LLSLHLPSFP HLCPGVISTP YTLTCPSNRS LPISWYCNGT RLHRLTASNL
    b-ugd29  VSSLLVCFPK LLSLHLPSFP PLCPGVISTP YTLTCPSNRF LPISWYCNGT RLHRITASNL
     b-ug52  VCSLAVCFPK LLSLHLPSFP HLCPGVISTP YTLTCPSTRS LSISWYCNGT RLHRITASNL
     b-g413  VCSLAVCFPK LLSLHLPSFP HLCPGVISTP YTLTCPSNRS LPISWYCNGT RLLRITASNP
     b-g482  VCSLAVCFPK LLSLHLPSFP HLCPGVLSTP YTLTCPSNRS LPISWYCNGT RLWRITASNL
    b-ugd13  VCSLLVCFPK LLSLHLPSFP HLCPGVVSTP YTLTCPSNRS LPISWYCNGT RLWRITASNL
     b-ugd3  VCSLLVCFPK LLSLHLPSFP HLCPGVVSTP YTLTCPSNRS LPISWYCNGT RLWRITASNL
    b-ugd15  VCSLLVCFPK LLSLHLPSFP HLCPGVVSTP YTLTCPSNRS LPISWYCNGT RLWRITASNL
     b-g71   VCSLLVCFPQ LLSFNLPSFP HLCPGVISTP YTLTCPSNRS LPISWYCNGT RLWRITASNL
     b-g91   LCNLVVCFPK LLSLHLPSFP HLCPGVISTP YTLTCPSNRS LPISWYCNGT QLLRITASNL
      a-bc1  VCSLAVCFRG LLSLSLLSSP NLCPGVISTP YTLTCLSNAS LPISWYCNNT RLLRLTERRV
     a-ema7  VCSLAVCFRG LLSLSLLSSP NLCPGVISTP YNLTCLSNAS LPISWYCNNT RLLRLTERRV
    a-uka13  VCSLAVCFRG LLSLSLLSSP NLCPGVISTP YNLTCLSNAS LPISWYCNNT RLFRLTERTL
    a-uka21  VCSLAVCFRG LLSLSLLSSP NLCPGVISTP YNLTCLSNAS LPISWYCNNT RLFRLTERTL
    a-ukma4  VCSLAVCFRG LLSLSLLSSP NLCPGVISTP YNLTCLSNAS LPISWYCNNT RLLRLTERTL
    a-bcblr  VCSLAVCFRG LLSLSLLSSP NLCPGVISTP YKLTCLSNAS LPISWYCNNT RLFRPTETTL
    a-ukb12  VCSLAVCFRG LLSLSLLSSP NLCPGVISTP YTLTCLSNAS LPIFWYCNNT RLFRLTKTIF
    a-ukc12  VCSLAVCFRG LLSLSLLSSP NLCPGVISTP YKLTCLSNAS LPISWYCNDT RLFRLTERTL
     a-iap1  VCSLAVCFRG LLSLSLLSSP NLCPGVISTP YTLTCLSDTS LPISWYCNNT RLLRLTTKTL
    a-ug111  VCSLAVCFRG LLSLSLLSSP NLCPGVISTP YNLTCLSNAS LPISWYCNNT RLFRLTEKTL
  c-k1-8-dem VCSLAVCFRG LLSLYVQSSP NLCPGVISTP YTLTCPSDAT LPISWYCNDT RLFRLTQQTL
    c-ugd23  VCSLAVCFRG LLSLYVQSSP NLCPGVISTP YTLTCPSDAT LPISWYCNDT RLLRLTQQTF
    a-ugd16  VCSLAVCFQG LLSLYLQSSP NLCPGVISTP YTLTCLSDAS LPISWYCNDT RLLRLTDQSF
    a-ugd18  VCSLAVCFQG LLSLYLQSSP NLCPGVISTP YTLTCLSDAS LPISWYCNDT RLLRLTDQSF
    a-ugd24  VCSLAVCFQG LLSLYLQSSP NLCPGVISTP YTLTCLSDAS LPISWYCNDT RLLRLTDQSF
    a-ug374  VCSLAVCFRG LLSLYLQSSP NLCPGVISTP YTLTCLSDAS LPISWYCNDT RLWRLTDQSF
     a-ugd4  VCSLAVCFRG LLSLYLQSSP NLCPGVISTP YTLTCLSDAS LPISWYCNDT RLWRLTDQSF
    a-ugd12  VCSLAVCFRG LLSLYLQSSP NLCPGVISTP YTLTCLSDAS LPISWYCNDT RLWRLTDQSF
    a-bcbl1  VCSLAVCFRG LLSLYVLSSP NLCPGVISTP YTLTCLSDSS LPISWYCNDT RLWRLTKPTL
     c-gk17  VCSLAVCFRG LLSLYLQSSP NMCPGVISTP YTLTCPSGTS LPTSWYCNDT RLFRLTQDTF
     c-iap3  VCSLAVCFRG LLSLYLQSSP NMCPGVISTP YTLTCPSNTS LPTSWYCNDT RLFRLTQDTF
     c-bbg1  VCSLAVCFRG LLSLSLQSSP NLCPGVISTP YTLTCPSNTS LPTSWYCNDT RLLRLTQQTF
    c-ukma8  VCSLAVCFRG LLSLSLQSSP NLCPGVISTP YTLTCPSNTS LPTSWYCNDT RLLRLTQQTF
      c-bc2  VCSLAVCFRG LLSLSLQSSP NLCPGVISTP YTLTCPSNTS LPTSWYCNDT RLLRLTQQTF
    c-ukma3  VCSLAVCFRG LLSLSLQSSP NLCPGVISTP YTLTCPSNTS LPTSWYCNDT RLLRLTQQTI
     c-ive1  VGSLAVCFQG LLSLSLQSSP NLCPGVILTP YTLTCPSNTS LPTSWYCNDT RLFRLTQQTL
    c-ukb22  VCSLAVCFRG LLSLSLQSSP NLCPRVISTP YTLTCPSNTS LPTSWYCNDT RLFRLTQQTL
     c-erla  VCSLAVCFRG LLSLSLQSSP NLCPGVISTP YTLTCPSNTS LPTSWYCNDT RLLRLTQQTF
    c-icam1  VCSLAVCFRG LLSLSLQSSP NLCPGVISTP YTLTCPSNTS LPTSWYCNDT RLLRLTQQTF
    c-ukma1  VCSLAVCFRG LLSLSLQSSP NLCPGVISTP YTLTCPSNTS LPTSWYCNDT RLFRLTQQTW
     c-iap2  VCSLAVCFRG LLSLSLQSSP NLCPGVISTT YTLTCPSNTS LPTSWYCNDT RLFRLTQQTF
     c-gk18  VCSLAVCFRG LLSLSLQSSP NLCPGVISTP YTLTCPSNTS LPTSWYCNDT RLLRVTQGTL
c-k1-43-ber  VCSLAVCFPG LLSLYVQSSP HLCPGVISTP YTLTCPSDAS LPISWYCNGT QLLRLTQRSV
   d1-tks10  VCSLAVCFPG LLSLHLSVSP QFCPAVLSTS YTLTCLSDAS LPISWYCNGT RLLRITGATL
    d2-zks3  VCNLAVCFPG LLSLHLPAFP PLCPGVLSTN YTLTCLSDAS LPISWYCNGT LLMRYHRTTL
  Consensus  -----V-F-- LLS------P --CP-V--T- Y-LTC-S--- L---WYCN-T -L-R------
```

```
            61                                                      120
     b-ugd2  IVSSLTGNFT CMTASGPTHS IWIEWHTTPV LQTLCAQPSN TVTCGQRVTL YCHTSSNNGT
    b-ugd21  TVSSLTGNFT CMTASGPTYS IWIEWYPTPV LQTLCAQPSN TVTCGQRVTL YCHTSSNNGT
     b-ugd1  TVSLLTCNFT CMTASGPTHS IWIEWYTTPV LQTLCAQPSN TVTCGQRVTL YCHTSSNNGT
    b-ugd26  TVSFLTCNFT CMTASGPTHS IWIEWYTTPV LQTLCAQPSN TVTCGQRVTL YCYTSSNNGT
    b-ugd19  TVASLTGNFT CMTASGPTHS IWIEWYTTPV LQTLCAQPSN TVTCGQRVTL YCHTSSNNGT
     b-ugd7  TVSSLTGNFT CMTASGPTHS IWIEWYTTPV LQTLCAQPSN TVTCGQRVTL YCHTSSNNGT
    b-ugd10  TVSSLTCNFT CMTTSGPTHS IWIQWYTQPV LQTLCAQPSN TVSCGQPVTL YCDTSSNNVT
   b-ukma24  TVSSLTCNFT CMTTSGPTHS IWIEWYTQPV LQTLCAQPSN TVSCGQPVTL YCDTSSNNVT
    b-ugd29  TVSSLTCNFT CMTTSGPTHS IWIEWYTQPV LQTLCTQPSN TVSCGQPVTL YCDTSSNNVT
     b-ug52  TVSSLTCNFT CMTTSGPTHS IWIEWYTQPV LQTLCAQPSN TVSCGQPVTL YCDTSSNNVT
     b-g413  TVCSLTCNFT CMTASGPTHS IWIEWYTQPV LQTLCAQPSN TVTCGQRVTL YCHTSSNNVT
     b-g482  TVSRLTCNFT CMTRSGPTHS IWIEWYTQPV LQTSCAQPSN TVTCGQRVTS YCHTSSNNVN
    b-ugd13  TVLSVTCNFT CMTTSGPTHS IWIEWYSQPV LQTLCAQPSN TVTCGQRVTL YCHTSSDNVT
     b-ugd3  TVLSVTCNFT CMTTSGPTHS IWIEWYSQPV LQTLCAQPSN TVTCGQRVTL YCHTSSDNVT
    b-ugd15  TVLSVTCNFT CMTTSGPTHS IWIEWYSQPV LQTLCAQPSN TVTCGQRVTL YCHTSSDNVT
     b-g71   TVSSLSCNFT CMTTSGPTHS IWIEWYTQPV LQTLCAQPSN TVTCGQRVTL YCHTSSNNVT
     b-g91   TVSAMTCNFT CMTTSGPTHS IWIEWYTEPV LQTLCAQPSN TVTCGQRVTL YCNTSSNNVT
     a-bc1   ILDTIACNFT CVEQSGHRQS IWITWRAQPV LQTLCAQPSN TVTCGQHVTL YCSTSGNNVT
     a-ema7  ILVTIACNFT CVEQSGHRQS IWITWHAQPV LQTLCAQPSN TVTCGQHVTL YCSTSGNNVT
    a-uka13  FSVTIACNFT CVEQSGHRQS IWITWHAQPV LQTLCAQPSN TVTCGQHVTL YCSTSGNNVT
    a-uka21  FSVTIACNFT CVEQSGHRQS IWITWHAQPV LQTLCAQPSN TVTCGQHVTL YCSTSGNNVT
    a-ukma4  FSVTIACNFT CVEQSGHRQS IWITWHAQPV LQTLCAQPSN TVTCGQHVTL YCSTSGNNVT
    a-bcblr  FPVTIACNFT CVEQSGHRQS IWITWHAQPV LQTLCAQPSN TVTCGQHVTL YCSTSGNNVT
    a-ukb12  STVTIACNFT CVEQSGHRQS IWITWHAQPV LQTLCAQPSN TVTCGQHVTL YCSTSGNNVT
    a-ukc12  FPVTIPCNFT CVEQSGHRQS IWVTWHAQPV LQTLCAQPSN TVTCGQHVTL YCSTSGNNVT
     a-iap1  YTFTMTCNFT CVEQSGHRQS IWISWHTQPV LQTLCAQPSN TVTCGQHVTL YCSTSGNNVT
    a-ug111  PGVNIACNFT CVEESGHPQN IWITWHAQPV LQTLCAQPSN TVTCGHHVTL YCSTSGNNVT
  c-k1-8-dem TFTNLTCNFT CVGQSGHRHS FWITWHAQPV LQTLCAQPSN TVTCAQHVTL YCSTSGNNVT
    c-ugd23  TFANLTCNFT CVGQSGYRHS VWITWHAQPV LQTLCAQPSN TVTCGQHVTL YCSTSGNNVT
    a-ugd16  TVATITCNFT CVEQSGHRQS IWITWNAQPV LQTLCAQPSN TVTCGQHVTL YCSTSGNNVT
    a-ugd18  TVATITCNFT CVEQSGHRQS IWITWNAQPV LQTLCAQPSN TVTCGQHVTL YCSTSGNNVT
    a-ugd24  TVATITCNFT CVEQSGHRQS IWITWNAQPV LQTLCAQPSN TVTCGQHVTL YCSTSGNNVT
    a-ug374  TVATITCNFT CVEQSGHRQS IWITWNAQPV LQTLCAQPSN TVACGQQVTL YCSTSGNNVT
     a-ugd4  TVATITCNFT CVEQSGHRQS IWITWNAQPV LQTLCAQPSN TITCGQHVTL YCSTSGNNVT
    a-ugd12  TVANITCNFT CVEQSGHRQS IWITWNAQPV LQTLCAQPSN TVACGQQVTL YCSTSGNNVT
    a-bcbl1  TIDIITCNFT CVEQSGHRQS IWITWHAQPV LQTLCAQPSN TVTCGQHVTL YCSTSGNNVT
     c-gk17  TVVNFMCNFS CVGQSGHRHS LWMAWYAPPV LQTLCGQPAN TVTCGQHVTL YCSTSGNNVT
     c-iap3  TVVNLMCNFS CVGQSGHRHS LWMAWYAPPV LQTLCGQPAN TVTCGQHVTL YCSTSGNNVT
     c-bbg1  TVVTLICNFS CVEQSGHRHS LWITWYPQPV LQTLCGQPSN TVTCGQHVTL YCSTSGNNVT
    c-ukma8  TVVTLICNFS CVGQSGHRHS LWITWYPQPV LQTLCGQPSN TVTCGQHVTL YCSTSGNNVT
     c-bc2   TVVALICNFS CVGQSGHRHS LWITWYPQPV LQTLCGQPSN TVTCGQHVTL YCSTSGNNVT
    c-ukma3  TVVTLICNFS CVGQSGHRHS LWITWYPQPV LQTLCGQPSN TVTCGQHVTL YCSTSGNNVT
     c-ive1  TVVTLICNFS CVGQSGHRHS LWITWYPQPV LQTLCGQPSN TVTCGQHVTL YCSTSGNNVT
    c-ukb22  TVVDLICNFS CVGQSGHRHS LWITWYPQPV LQTLCGQSSN TVTCGQHVTL YCSTSGNNVT
     c-erla  TVDTLICNFS CVGQSGHRHS LWITWYPQPV LQTLCGQPSN TVTCGQHVTL YCSTSGNNVT
    c-icam1  TVVALICNFS CVGQSGHRHS LWITWYREPV LQTLCGQPSN TVTCGQHVTL YCSTSGNNVT
    c-ukma1  TVHTLICNFS CVGQSGHRHS LWITWYPQPV LQTLCGQPSN TVTCGQHVTL YCSTSGNNVT
     c-iap2  TVNTLICNFT CVGQSGHRHS LWITWYAQPV LQTFCGQPSN TVTCGQHVTL YCSTSGNNVT
     c-gk18  TVDTLICNFS CVGQSGHRYS LWITWYAQPV LQTFCGQPSN TVTCGQHVTL YCSTSGNNVT
c-k1-43-ber  PVSTLTCNFT CVGQSGPSHS IWITWYAKPV LQTLCAQPSN TVTCGQHVPL YCSTSGNNVT
   d1-tks10  TIPSLTGNFT CVDHSGLSHS IWIQRYPPPV LQTLCAQPST TVTCGQRVSL HCSTSANTVT
    d2-zks3  TLMNLAANWT CVNQSGISHS IWIQWYTEPV LQTLCAQPSN TVTCGQRVTL HCSTSASTVF
  Consensus  -------N-- C---SG---- -W------PV LQT-C-Q--- T--C---V-- -C-TS-----
```

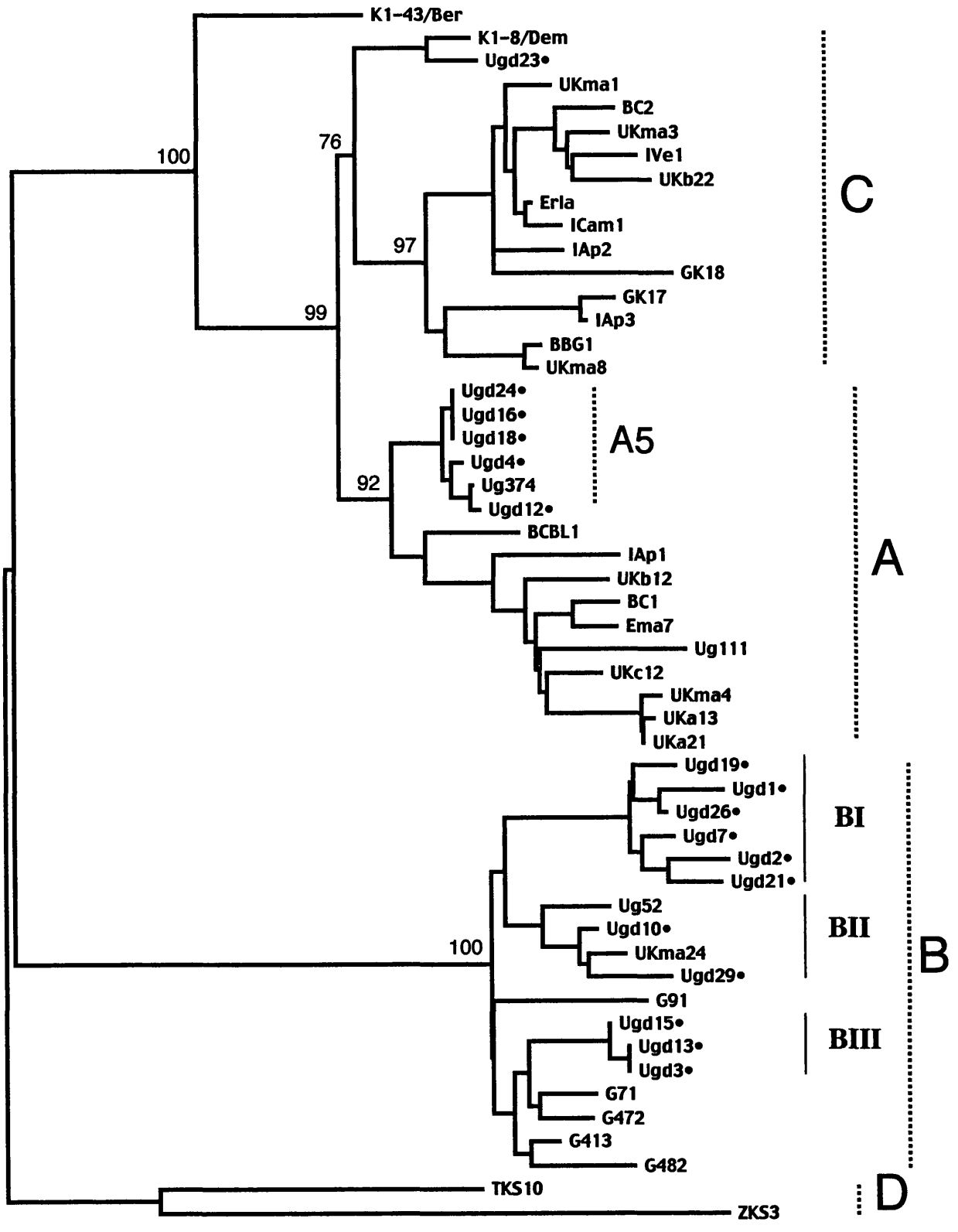
```
            121                                                                 180
     b-ugd2 IWRLQNGHNQ TVSQTKYYNF TMMNQTQGCY ICCDGLSSHL SNRICFWASC INITPETPTV
    b-ugd21 IWRLQNGHNQ TVSQTKYYNF TMMNQTQGCY ICCDGLSSRL SNRICFWASC INITPETPTV
     b-ugd1 IWRLQNGHNQ TVSQTKYYNF TMMNQTQGCY LCSDGLSSRL SNRICFWASC INITPETPTV
    b-ugd26 IWRLQNGHNQ TVSQTKYYNF TMMNQTQGCY LCSDGLSSRL SNRICFWASC INITPETPTV
    b-ugd19 IWRLQKGHNQ TVSQTKYYNF TMMNQTEGCY LCSDGLSSRL SNRICFWASC INITPETPTV
     b-ugd7 IWRLQNGHNQ TVSQTKYYNF TMMNQTQGCY FCFDGLSSRL SNRICFWASC INITPETPTV
    b-ugd10 IWRLQNGHNQ TVSQTKYYNF TMMNQTQGCY TCSNGLSSRL SNRICFWASC INITPETPTV
   b-ukma24 IWRLQNGHNQ TVSQTKYYNF TMMNQTQGCY TCSNGLSSRL SNRICFWASC INITPETPTV
    b-ugd29 IWRLQTGHNQ TVSQTKYYNF TMMNQTQGCY ACSNGLSSRL SNRICFWASC INVTPETPTV
     b-ug52 IWRLQNGHNQ TVSQTRYYNF TMMNQTRGCY VCSDGLSSRL SNRICFWASC INITPETPTV
     b-g413 IWRLQNGHNQ TVSQTKYYNF TMMNQTQGCY TCSNGLSSRL SNRICFWASC INITPETPTV
     b-g482 IWRLQNGHNQ TVSQTKYYNF TMMNQTQGCY TCSNGLSSRL SNRICFWASC INITPETPTV
    b-ugd13 IWRLQNGHNQ TVSQTKYYNF TMMDQTQGCY TCSNGLSSRL SNRICFWASC INITPETPTV
     b-ugd3 IWRLQNGHNQ TVSQTKYYNF TMMDQTQGCY TCSNGLSSRL SNRICFWASC INITPETPTV
    b-ugd15 IWRLQNGHNQ TVSQTKYYNF TMMNQTQGCY TCSNGLSSRL SNRICFWASC INITPETPTV
     b-g71 IWRLQNGHNQ TVSQTKYYNF TMMNQTQGCY TCSNGLSSRL SNRICFWASC INITPETPTV
     b-g91 IWRLQNGHNQ TVSQTKYYNF TMMNQTQGCY TCSNGLSSRL SNRICFWASC INLTPETPTL
     a-bc1 VWHLPNGRNE TVSQTKYYNF TLMSQTEGCY TCSNGLSSRL SNRICFWARC ANITPETHTV
    a-ema7 VWHLPNGRNE TVSQTKYYNF TLMNQTEGCY TCSNGLSSRL SNRICFWARC DNITPETHTV
   a-uka13 VWHLPNGRNE TVSQTKYYNF TLMNQTEGCY TCSNGLSSRL SNRICFWARC ANLTPETHTV
   a-uka21 VWHLPNGRNE TVSQTKYYNF TLMNQTEGCY TCSNGLSSRL SNRICFWARC ANITPETHTV
   a-ukma4 VWHLPNGRNE TVSQTKYYNF TLMNQTEGCY TCSNGLSSRL SNRICFWARC ANITPETHTV
   a-bcblr VWHLPNGRNE TVSQTKYYNF TLMNQTEGCY TCSNGLSSRL SNRICFWARC ANITPETHTV
   a-ukb12 VWHLPNGRNE TVSQTKYYNF TLMNQTEGCY TCSNGLSSRL SNRICFWARC ANITPETHTV
   a-ukc12 VWHLPNGRNE TVSQTKYYNF TLMNQTEGCY TCSNGLSSRL SNRICFWARC ANITPETQTV
    a-iap1 VWHLPNGRNE TVSQTKYYNF TLMNQTEGCY TCSNGLSSRL SNRICFWARC ANITPETHTV
   a-ug111 VWHLPNGLNE TVSQTKYYNF TPMNQTEGCY TCSNGLSSRL SNRICFWARC ANITPETHTV
c-k1-8-dem VWHLPNGRNE TVSQTKYYNF TLMNQTEGCY ACSNGLSSRL SNRICFWARC ANITPETHTV
   c-ugd23 VWHLPNGRNE TVSQTKYYNF TLMNQTEGCY GCSNGLSSRL SNRICFWARC ANITPETHTV
   a-ugd16 VWHLPNGRNE TVSQTKYYNF TLMNQTEGCY TCSNGLSSRL SNRICFWARC ANITPETHTV
   a-ugd18 VWHLPNGRNE TVSQTKYYNF TLMNQTEGCY TCSNGLSSRL SNRICFWARC ANITPETHTV
   a-ugd24 VWHLPNGRNE TVSQTKYYNF TLMNQTEGCY TCSNGLSSRL SNRICFWARC ANITPETHTV
   a-ug374 VWHLPNGRNE TVSQTKYYNF TLMNQTEGCY TCSNGLSSRL SNRICFWARC ANITPETHTV
    a-ugd4 VWHLPNGRNE TVSQTKYYNF TLMNQTEGCY TCSNGLSSRL SNRICFWARC ANITPETHTV
   a-ugd12 VWHLPNGRNE TVSQTKYYNF TLMNQTEGCY TCSNGLSSRL SNRICFWARC ANITPETHTV
   a-bcbl1 VWHLPNGRNE TVSQTKYYNF TLMNQTEGCY TCSNGLSSRL SNRICFWARC ANITPETHTV
    c-gk17 VWHLPNGRNE TVSQTKYYNF TLMNQTEGCY ACSNGLSSGL SNRLCFWARC ANITPETHTV
    c-iap3 VWHLPNGRNE TVSQTKYYNF TLMNQTEGCY ACSNGLSSGL SNRLCFWARC ANITPETHTV
    c-bbg1 VWHLPNGRNE TVSQTKYYNF TLMDQTEGCY ACSDGLSSRL SNRLCFWARC ANITPETHTV
   c-ukma8 VWHLPNGRNE TVSQTKYYNF TLMDQTEGCY ACSDGLSSRL SNRLCFWARC ANITPETHTV
     c-bc2 VWHLPNGRNE TVSQTKYYNF TLMDQSEGCY ACSDGLSSRL SNPLCFSARC ANITLETDTV
   c-ukma3 VWHLPNGRNE TVSQTKYYNF TLMDQTEGCY ACSDGLSSRL SNRLCFSARC ANITLETNTV
    c-ive1 VWHLPNGRNE TVSQTKYYNF TLMDQTEGCY ACSDGLSSRL SNRLCFSARC ANITLETHTV
   c-ukb22 VWHLPNGRNE TVSQTKYYNF TLMDQTEGCY ACSDGLSSRL SNRLCFSARC ANITLETHTV
    c-erla VWHLPNGRNE TVSQTKYYNF TLMNQTEGCY ACSNGLSSRL SNRLCFSARC ANLTPETHTV
   c-icam1 VWHLPNGRNE TVSQTKYYNF TLMNQTEGCY ACSNGLSSRL SNRLCFSARC ANLTPETHTV
   c-ukma1 VWHLPNGRNE TVSQTKYYNF TLMNQTEGCY ACSNGLSSRL SNRLCFSARC ANITPETHTV
    c-iap2 VWHLPNGRNE TVSQTKYYNF TLMNQTEGCY ACSNGLSSRL SNRLCFSARC ANRTPETHTV
    c-gk18 VWHLPNGQNE TVSQTKYYNF TLMNQTEGCY ACSNGLSSRL SNRLCFSARC ANITPETHTV
c-k1-43-ber IWHLQNGRNQ TVSQTNYYNF TVMNQTEGCY ACSNGLSSRL SNPICFWPRC ANITPETHTV
   d1-tks10 IWRLQNEGHQ PVSQTYYYNF TLMNQTQGCY TCSNGLSSRM SNRICFWAPC ANITPETDTV
    d2-zks3 IWRLHRGGNQ TVSQTNYYNF TLMNQTNGCY ACSSGLSSRL SNLICFWAHC ANISLETSTV
  Consensus -W-L------ -VSQT-YYNF T-M-Q--GCY -C--GLSS-- SN--CF---C -N---ET-T-
```

```
            181                                                                240
     b-ugd2  SASSTTAIK. ..........  ..MSRTNGLL KIIPATTHAA VAVEEVKSTN PHIQVPFLVF
    b-ugd21  SASSTTAIK. ..........  ..MSRTNGLL KIIPATTHAA VAVEEVKSTN PHIQVPFLVF
     b-ugd1  SASSTTAFK. ..........  ..MSRTNGLL KIIPATTHAA VAVEEVKSTN PHIQVPFLVF
    b-ugd26  SASSTTAFK. ..........  ..MSRTNGLL KIIPATTHAA VAVEEVKSTN PHIQVPFLVF
    b-ugd19  SASSTTAFK. ..........  ..MSRTNGLL KIIPATTHAA VAVEEVKSTN PHIQVPFLVF
     b-ugd7  SASSTTAFK. ..........  ..MSRTNGLL KIIPATTHAA VAVEEVKSTN PHIQVPFLVF
    b-ugd10  SASSTMAVK. ..........  ..LLRTNGLL KIIPATTHAA VAVEEVKSTN THIQVPFLVF
   b-ukma24  SASSTMAVK. ..........  ..VLRTNGLL KIIPATTHAA VAVEEVKSTN THIQVPFLVF
    b-ugd29  SASSPMAVK. ..........  ..VLRTNGLL KIIPATTHAA VAVEEVKSTN THIQVPFLVF
     b-ug52  SASSTTAFK. ..........  ..MSRTNGLL KIIPATTHAA VAVEEVKSTN PHIQVPFLVF
     b-g413  SSSSTTAFE. ..........  ..TLRTNGLL KIIPATTHAA VAVEEVKSTN PHIQVPFLVF
     b-g482  SSSSTTAFK. ..........  ..TLRTNGLL KIIPATTHAA VAVEEVKSTN PHIQVPFLVF
    b-ugd13  SASSTTAFK. ..........  ..TLTTNGLL KIIPATTHAA VAVEEVKSTN PHIQVPFLVF
     b-ugd3  SASSTTAFK. ..........  ..TLTTNGLL KIIPATTHAA VAVEEVKSTN PHIQVPFLVF
    b-ugd15  SASSTTAFK. ..........  ..TLTTNGLL KIIPATTHAA VAVEEVKSTN PHIQVPFLVF
     b-g71   SASSTTAFK. ..........  ..KLRNNGLL KIIPATTHAA VAVEEVKSTN PHIQVPFLVF
     b-g91   SASSTTAFK. ..........  ..TLRTNGLL KIIPATTHAA VAVEEVKSTN PHIQVPFLVF
      a-bc1  SVSSTTGFR. ..........  ..TLSTNSLV KIIHATTRDV VVVKEAKSTH FHIEVHFLVF
     a-ema7  SVSSTTGFR. ..........  ..TLSTNSLV NIIHATTHDV VVVKEAKSTH FHIELHFLVF
    a-uka13  SVSSTTGFR. ..........  ..TFSTNRLV NIIHATTHDV VVLKEAKSTR FHIELHFLVF
    a-uka21  SVSSTTGFR. ..........  ..TFSTNRLV NIIHATTHDV VVLKEAKSTR FHIELHFLVF
    a-ukma4  SVSSTTGFR. ..........  ..TFSTNRLV NIIHATTHDV VVLKEAKSIR FHIELHFLVF
    a-bcblr  SVSSTTGFR. ..........  ..TFSTNSLV NIIHATTHDV VVVKEAKSTH FHIELHFLVF
    a-ukb12  SVSSTTGFR. ..........  ..TFSTNSLV NIIHATTHDV VAVKEAKSTH FHIELHFLVF
    a-ukc12  SVSSTTGFR. ..........  ..TFSTNSLV NII....HDV VVVKEAKSTH FHIELHFLVF
     a-iap1  SVSSTTGFR. ..........  ..TFSTNSLV NIIHATTHDA VVVKEAKSTR FHIEVHFLVF
    a-ug111  SVSSTTGFR. ..........  ..TFSTNSLV NIIHATTHNV VVVKEAKSTH FHIELHFLVF
c-k1-8-dem   SVSSTTGFG. ..........  ..TFSTHRLV NRIHATTHDV VVVKEAKSTN PHIEVPFLVF
    c-ugd23  SVSSTTGFG. ..........  ..TLSTHSLV NRIHATTHDV VVVKEAKSTN PHIEVPFLVF
    a-ugd16  SVSSTTGFR. ..........  ..TVSTNSLV NIIHATNHDV VVVKEAKSTN PHIEVPFLVF
    a-ugd18  SVSSTTGFR. ..........  ..TVSTNSLV NIIHATNHDV VVVKEAKSTN PHIEVPFLVF
    a-ugd24  SVSSTTGFR. ..........  ..TVSTNSLV NIIHATTHDV VVVKEAKSTN PHIEVPFLVF
    a-ug374  SVSSTTGFR. ..........  ..TFSTNSLV NIIHATTHDV VVVKEAKSTN PHIEVPFLVF
     a-ugd4  SVSSTTGFR. ..........  ..NLSTHSLV NIIHATTHHV VVVKEAKSTN PHIEVPFLVF
    a-ugd12  SVSST..... ..........  .......... .IIHATTHDV VVVKEAKSTN PHIEVPFLVF
    a-bcbl1  SVSSTTGFR. ..........  ..TFSTNSLV NIIHATTHKV VVVKEAKSTN SHIEVHFLVF
     c-gk17  SVSSTTGFR. ..........  ..TFSTHS..  ...HTTTHAV VVVKEAKFTN PHIEVPFLVF
     c-iap3  SVSSTTGFR. ..........  ..TFSTHS..  ...HTTTHAV VVVKEAKFTN PHIEVPFLVF
     c-bbg1  SVSSTTGFR. ..........  ..TFSTNS..  ...HATKHDV VVVKEAKFTN PHIEVPFLVF
    c-ukma8  SVSSTTGFR. ..........  ..TFSTNS..  ...HATKHDV VVVKEAKFTN PHIEVPFLVF
      c-bc2  SVSSTTGFR. ..........  ..TFSTNS..  ...AATTHDV LVMKEAKSTN LHIQVHFIVF
    c-ukma3  SVSSTTGFR. ..........  ..TFSTNS..  ...DATTHDI LVMKEAKSTN LHIQVHFLVF
     c-ive1  SVSSTTGFR. ..........  ..TFSTNS..  ...HATTHDE LVMKEAKSTN LHIQVHFLVF
    c-ukb22  SVSSTTGFR. ..........  ..TFSTNS..  ...HATTHDV LVMKEAKSTN LPIQVHFFVF
     c-erla  SVSSTTGFR. ..........  ..TFSTNR..  ...RGTTLDV LVMKEAKSTN LHIQVHFLVF
    c-icam1  SVSSTTGFT. ..........  ..TFSTNR..  ...RGTTLDV LVMKEAKSTN LHIQVHFLVF
    c-ukma1  SVSSTTGFR. ..........  ..TFSTNR..  ...HAATHDI LVMKEAKSTN LHIQVHFLVF
     c-iap2  SVSSTTGFR. ..........  ..TFSTNS..  ...HTTAHNV LVMKEAKSTN LHIQVHSLVF
     c-gk18  SVSSTTGFR. ..........  ..TF......  ....ATAPTL FVMKEVKSTY LYIQEHLLVF
c-k1-43-ber  SVSSTTGFR. ..........  ..TFSTNSIV NITHATTRAV VVVKEAKSTN PHIEVPFLVF
    d1-tks10 SVRSTTGFKT HTVSVRSTTG FTTFSTNRLV NIIPATTHAV VVVEKVKSLH PHIEVPFLVF
    d2-zks3  SVSSTTGFK. ..........  ..TFSTNRLV NIIPATTHAV VVVEELKSRN PYIKVHFLIL
   Consensus S--S------ ----------  ---------- ----------  ------K---  --I-------
```

```
            241                                                        293
     b-ugd2  MTLVALIGTM CGILGTLIFT HCQKKSDSSK TGQQQLRDYY SLDYFHTEEY TQP
    b-ugd21  MTLVALIGTM CGILGTLIFT HCQKKSDSSK TGQQQLRDYY SLDYFHTEEY TQP
     b-ugd1  MTLVALIGTM CGILGTLIFT HCQKKSDSSK TGQQQLRDYY SLDYFHTEEY TQP
    b-ugd26  MTLVALIGTM CGILGTLIFT HCQKKSDSSK TGQQQLRDYY SLDYFHTEEY TQP
    b-ugd19  MTLVALIGTM CGILGTLIFT HCQKKSDSSK TGQQQLRDYY SLDYFHTEEY TQP
     b-ugd7  MTLVALIGTM CGILGTLIFT HCQKKSDSSK TGQQQLRDYY SLDYFHTEEY TQP
    b-ugd10  MTLVALIGTM CGILGTLIFT HCQKKSDSSK TGQQQLRDYY SLDYFHTEDY TQP
   b-ukma24  MTLVALIGTM CGILGTLIFT HCQKKSDSSK TGQQQLRDYY SLDYFHTEDY TQP
    b-ugd29  MTLVALIGTM CGILGTLIFT HCQKKSDSSK TGQQQLRDYY SLDYFHTEDY TQP
     b-ug52  MTLVALIGTM CGILGTLIFT HCQKKSDSSK TGQQQLRDYY SLDYFHTEDY TQP
     b-g413  MTLVALIGTM CGILGTVIFT HCQKKSDSSK TGQQQLRDYY SLDYFHTEDY TQP
     b-g482  ITLVALIGTM CGILGAVIFT HCQKKSDSSK TGQQQLRDYY SLDYFHTEDY TQP
    b-ugd13  MTLVALIGTM CGILGTVIFT HCQKKSDSSK TGQQQLRGYY SLDYFHTEEY TQP
     b-ugd3  MTLVALIGTM CGILGTVIFT HCQKKSDSSK TGQQQLRGYY SLDYFHTEEY TQP
    b-ugd15  MTLVALIGTM CGILGTVIFT HCQKKSDSSK TGQQQLRDYY SLDYFHTEEY TQP
      b-g71  MTLVALIGTM CGILGTVIFT HCQKKSDSSK TGQQQLPDYY SLDYFHTEEY TQP
      b-g91  MTLVALIGTM CGIFGTLIFA HCQKQSDSSN TGQQQLRDYY SLDYFHTEEY TQP
      a-bc1  MTLVALIGTM CGILGTIIFA HCQKQRDSNK TVPQQLQDYY SLHDLCTEDY TQP
     a-ema7  MTLVALIGTV CGILGTIIFA HCQKQRDSNK TVPQQLQDYY SLHDLCTEDY TQP
    a-uka13  MTLVALIGTM CGIVGTIIFF HCQKQRDSNK TVPQQLQDYY SLHDLCTEDY TQP
    a-uka21  MTLVALIGTM CGIVGTIIFF HCQKQRDSNK TVPQQLQDYY SLHDLCTEDY TQP
    a-ukma4  MTLVALIGTM CGIVGTIIFF HCQKQRDSNK TVPQQLQDYY SLHDLCTEDY TQP
    a-bcblr  MTLVALIGTM CGILGTIIFA HCQKQRDSNK TVPQQLQDYY SLHDLCTEDY TQP
    a-ukb12  MTLVALIGTM CGILGTIIFA HCQKQRDSNK TVPQQLQDYY SLHDLCTEDY TQP
    a-ukc12  MTLVALIGTM CGILGTIIFA HCQKQRDSNK TVPQQLQDYH SLHDLCTEDY TQP
     a-iap1  MTLVALIGTM CGILGTIIFA HCQKQRDSNK TVPQQLQDYY SLHDLCTEDY TQP
    a-ug111  MTLVALIGTM CGILGTIIFA HCRKQRDSNK PVPQQLQDYY SLHDLCTEDY TQP
c-k1-8-dem  MTLVALIGTM CGILGTIIFA HCQKQSDSKK TVPQQLRDYY SLHDFCTEDY TQP
    c-ugd23  MTLVALIGTM CGILGTIIFA HCQKQSDSNK TVPQQLRDYY SLHDFCTEDY TQP
    a-ugd16  MTLVALIGTM CGILGTIIFA HCQKQRDSNK TVPQQLQDYY SLHDLCTEDY TQP
    a-ugd18  MTLVALIGTM CGILGTIIFA HCQKQRDSNK TVPQQLQDYY SLHDLCTEDY TQP
    a-ugd24  MTLVALIGTM CGILGTIIFA HCQKQRDSNK TVPQQLQDYY SLHDLCTEDY TQP
    a-ug374  MTLVALIGTM CGILGTIIFA HCQKQRDSNK TVPQQLQDYY SLHDLCTEDY TQP
     a-ugd4  MTLVALIGTM CGILGTIIFA HCQKQRDSNK TVPQQLQDYY SLHDLCTEDY TQP
    a-ugd12  MTLVALIGTM CGILGTIIFA HCQKQRDSNK TVPQQLQDYY SLHDLCTEDY TQP
    a-bcbl1  MTLVALIGTM CGILGTIIFA HCQKQRDSNK TVPQQLQDYY SLHDLCTEDY TQP
     c-gk17  MTLVALIGTM CGILGTIIFA HCQKQSDSNK TVPQQLRDYY SLHDLCTEDY TQP
     c-iap3  MTLVALIGTM CGILGTIIFA HCQKQSDSNK TVPQQLRDYY SLHDLCTEDY TQP
     c-bbg1  MTLVALIGTM CGILGTIIFA HCQKQSDSNK TVPQQLRDYY SLHDLCTEDY TQP
    c-ukma8  MTLVALIGTM CGILGTIIFA HCQKQSDSNK TVPQQLRDYY SLHDLCTEDY TQP
      c-bc2  MTLVALIGTM CGILGTIIFA HCQKQRDSNK TVPQQLQDYY SLHDLCTEDY TQP
    c-ukma3  MTLVALIGTM CGILGTIIFA HCQKQSDSNK TVPQQLRDYY SLHDLCTEDY TQP
     c-ive1  MTLVALIGTM CGILGTIIFA HCQKQSDSNK TVPQQLRDYY SLHDLCTEDY TQP
    c-ukb22  MTLVALIGTM CGILGTIIFA HCQKQSDSNK TVPQQLRDYY SLHDLCTEDY TQP
     c-erla  MTLVALIGTM CGILGTIIFA HCQKQRDSNK TVPQQLQDYY SLHDLCTEDY TQP
    c-icam1  MTLVALIGTM CGILGTIIFA HCQKQRDSNK TVPQQLQDYY SLHDLCTEDY TQP
    c-ukma1  MTLVALIGTM CGILGTIIFA HCQKQRDSNK TVPQQLQDYY SLHDLCTEDY TQP
     c-iap2  MTLVALIGTM CGILGTIIFA HCQKQRDSNK TVPQQLQDYY SLHDLCTEDY TQP
     c-gk18  MTLVALIGTM CGILGTIIFA HCQKQRDSNK TVPQQLQDYY SLHDLCTEDY TQP
c-k1-43-ber  MTLVALIGTM CGILGTIIFA HCQKQSDSNK TVPQQLRDYY SLHDLCTEDY TQP
   d1-tks10  MTLVALIGTM CGILGTIIFA HCQKQSDSNK TVQQQLRDYY SLHDFNTEDY TQP
    d2-zks3  MTLVALIGTM CGILGTIIFA RCQKQSDSNK TVPQQLRDYY SLHDFITEDY MQP
  Consensus  -TLVALIGT- CGI-G--IF- -C-K--DS-- ---QQL--Y- SL----TE-Y -QP
```

Fig. 4.6. HHV-8 K1 phylogenetic tree.

A neighbour joining tree comprising 52 almost complete K1 DNA sequences (840 bp, codons 6-285), 17 from this study (highlighted by black dots) and 35 from previous studies (Cook et al., 1999; Zong et al., 1999; Lacoste et al., 2000a). It was rooted at the midpoint between the sum of the longest branches (Ugd2 and ZKS3). Four subtypes (A-D), A5 variants and the three B clusters (BI, BII and BIII) to which the Ugandan B subtype strains belong are indicated. Selected bootstrap values are included. The bar at the bottom of the tree shows the divergence scale (substitutions/site). The tree was constructed by Duncan McGeoch and formatted in Treeview by Andrew Davison.


K1-43/Ber
K1-8/Dem
Ugd23•
UKma1
BC2
UKma3
IVe1
UKb22
Erla
ICam1
IAp2
GK18
GK17
IAp3
BBG1
UKma8
Ugd24•
Ugd16•
Ugd18•
Ugd4•
Ug374
Ugd12•
BCBL1
IAp1
UKb12
BC1
Ema7
Ug111
UKc12
UKma4
UKa13
UKa21
Ugd19•
Ugd1•
Ugd26•
Ugd7•
Ugd2•
Ugd21•
Ug52
Ugd10•
UKma24
Ugd29•
G91
Ugd15•
Ugd13•
Ugd3•
G71
G472
G413
G482
TKS10
ZKS3
100
76
97
99
92
100
C
A5
A
BI
BII
BIII
B
D
0.01

The majority (11) of the Ugandan samples cluster with the B subtype, five (Ugd4, Ugd12, Ugd16, Ugd18 and Ugd24) with the A subtype and one (Ugd23) with the C subtype. The Ugandan B subtype strains fall into three clusters that have been designated as BI, BII and BIII (Fig. 4.6) for ease of reference. All the Ugandan A strains cluster closely with the previously identified A5 variant, Ug374 (Cook et al., 1999). Ugd23 clusters closely with K1-8/Dem (Lacoste et al., 2000a), and both branch early in the C lineage. The relatively low bootstrap value (76) for this node implies that inclusion of these strains in the C subtype is tentative.

Visual inspection of the K1 protein sequence alignment (Fig. 4.5B) reveals that both Ugd23 and K1-8/Dem do not have the characteristic 5-aa deletion of subtype C strains. Instead, they have a sequence that differs from that in subtype A strains by only one amino acid residue. In addition to unique differences (three each), characteristics of the A and C subtypes are present throughout their sequences. These data argue against recombination between an A and C subtype to generate these strains. The Ugd23 and K1-8/Dem K1 protein sequences differ by nine amino acid residues.

4.6 VARIABILITY WITHIN THE UGANDAN K1 SEQUENCES

DNA and amino acid alignments comprising the 17 Ugandan K1 sequences were made (Fig. 4.7). The results show that all strains have 289 amino acids in the complete K1 protein, except Ugd12, which has a unique deletion of 13 residues (aa 191-203). This deletion has not been reported in any other strain.

In order to identify the least and most widely diverged sequences within each subtype, distances between pairs of K1 nucleotide and amino acid sequences were determined. The values are shown in Table 4.3. The range of pairwise

Fig. 4.7. Alignments of K1 DNA and amino acid sequences of 17 HHV-8 strains.

The alignments include almost complete K1 DNA (A) or amino acid (B) sequences of strains analysed in this study. Prefixes on sequence names denote K1 subtypes. The alignments run in the orientation of the genomic sequence, nucleotide 1 and 855 corresponding to 120 and 974, respectively in the BC-1 genome. Fifteen base pairs were omitted at the 5'-end of the DNA sequences, while 5 amino acids at the N-terminus of the protein sequences were omitted; thus positions 1 and 284 are equivalent to aa 6 and 289 (end of the predicted protein), respectively. The two hypervariable regions (VR1, VR2) are indicated. Amino acids that replaced two cysteine residues in some of the Ugandan strains are highlighted in red. Key amino acids in the predicted ITAM are highlighted by asterisks. TM, transmembrane domain; CYT, cytoplasmic domain.

A

```
          1                                                          60
 b-ugd1   CTCTGCTGTT TGGTGGTTTG GTTTCCAAAA CTATTGAGCC TTCATCTGCC ATCGTTTCCA
b-ugd26   CTCTGCAGTT TGGTGGTTTG GTTTCCAAAA CTATTGAGCC TTCATCTGCC ATCGTTTCCA
b-ugd19   CTCTGCAGTT TGGTGGTTTG GTTTCCAAAA CTATTGAGCC TTCATCTGCC ATCGTTTCCA
 b-ugd7   CTCTGCAGTT TGGTGGTTTG CTTTCCAAAA CTATTGAGCC TTCATCTGCC ATCGTTTCCA
 b-ugd2   CTCTGCAGTT TGGTGGTTTG GTTTCCAAAA CTATTGAGCC TTCATCTGCC ATCGTTTCCA
b-ugd21   CTCTGCAGTT TGGTGGTTTG GTTTCCACAA CTATTGAGCC TTCATCTGCC ATCGTTTCCA
b-ugd13   GTCTGCAGTT TGCTGGTTTG CTTTCCAAAA CTATTGAGCC TTCATCTGCC ATCGTTTCCA
 b-ugd3   GTCTGCAGTT TGCTGGTTTG CTTTCCAAAA CTATTGAGCC TTCATCTGCC ATCGTTTCCA
b-ugd15   GTCTGCAGTT TGCTGGTTTG CTTTCCAAAA CTATTGAGCC TTCATCTGCC ATCGTTTCCA
b-ugd10   GTCTGCAGTT TGCTGGTTTG CTTTCCAAAA CTATTGAGCC TTCATCTGCC ATCGTTTCCA
b-ugd29   GTCTCCAGTT TGCTGGTTTG CTTTCCAAAA CTATTGAGCC TTCATCTGCC ATCGTTTCCA
a-ugd16   GTCTGCAGTC TGGCGGTTTG CTTTCAAGGA CTATTAAGCC TTTATCTGCA ATCGTCTCCA
a-ugd18   GTCTGCAGTC TGGCGGTTTG CTTTCAAGGA CTATTAAGCC TTTATCTGCA ATCGTCTCCA
a-ugd24   GTCTGCAGTC TGGCGGTTTG CTTTCAAGGA CTATTAAGCC TTTATCTGCA ATCGTCTCCA
 a-ugd4   GTCTGCAGTC TGGCGGTTTG CTTTCGAGGA CTATTAAGCC TTTATCTGCA ATCGTCTCCA
c-ugd23   GTCTGCAGTC TGGCGGTTTG CTTTCGAGGA CTATTAAGCC TTTATGTGCA ATCGTCTCCA
a-ugd12   GTCTGCAGTC TGGCGGTTTG CTTTCGAGGA CTATTAAGCC TTTATCTGCA ATCGTCTCCA
   Cons   -TCT-C-GT- TG--GGTTTG -TTTC-A--A CTATT-AGCC TT-AT-TGC- ATCGT-TCCA

          61                                                         120
 b-ugd1   CATTTGTGCC CTGGAGTGCT TTTCACGCCT TACACGTTGA CTTGTCCGTC TAACAGATCC
b-ugd26   CATTTGTGCC CTGGAGTGCT TTTCACGCCT TACACGTTGA CTTGTCCGTC TAACAGATCC
b-ugd19   CATTTGTGCC CTGGAGTGCT TTTCACGCCT TACACGTTGA CTTGTCCGTC TAACAGATCC
 b-ugd7   CATTTGTGCC CTGGAGTGCT TTTCACGCCT TACACGTTGA CTTGTCCGTC TAACAGATCC
 b-ugd2   CATTTGTGCC CTGGAGTGCA TTTCACGCCT TACACGTTGA CTTGTCCGTC TAACAGATCC
b-ugd21   CATTTGTGCC CTGGAGTGCA TTTCACGTCT TACACGTTGA CTTGTCCGTC TAACAGATCC
b-ugd13   CATTTGTGCC CTGGAGTGGT TTCCACGCCT TACACGTTGA CCTGTCCGTC TAACAGATCC
 b-ugd3   CATTTGTGCC CTGGAGTGGT TTCCACGCCT TACACGTTGA CCTGTCCGTC TAACAGATCC
b-ugd15   CATTTGTGCC CTGGAGTGGT TTCCACGCCT TACACGTTGA CCTGTCCGTC TAACAGATCC
b-ugd10   CATTTGTGCC CTGGAGTGAT TTCCACGCCT TACACGTTGA CCTGTCCGTC TAACAGATCC
b-ugd29   CCTTTGTGCC CTGGAGTGAT TTCCACGCCT TACACGTTGA CCTGTCCGTC TAACAGATTC
a-ugd16   AATTTGTGCC CTGGAGTGAT TTCAACGCCT TACACGTTGA CCTGTCTGTC TGATGCATCC
a-ugd18   AATTTGTGCC CTGGAGTGAT TTCAACGCCT TACACGTTGA CCTGTCTGTC TGATGCATCC
a-ugd24   AATTTGTGCC CTGGAGTGAT TTCAACGCCT TACACGTTGA CCTGTCTGTC TGATGCATCC
 a-ugd4   AATTTGTGCC CTGGAGTGAT TTCAACGCCT TACACGTTGA CCTGTCTGTC TGATGCATCC
c-ugd23   AATTTGTGCC CTGGAGTGAT TTCAACGCCT TACACGTTGA CCTGTCCGTC TGATGCAACC
a-ugd12   AATTTGTGCC CTGGAGTGAT TTCAACGCCT TACACGTTGA CCTGTCTGTC TGATGCATCC
   Cons   --TTTGTGCC CTGGAGTG-- TT--ACG-CT TACACGTTGA C-TGTC-GTC T-A---A--C

          121                                                        180
 b-ugd1   TTGCCAACAT CCTGGTATTG CAACGGGACT CAGCTTCGGC GAATACGGGG GTCTAACCTA
b-ugd26   TTGCCAATAT CCTGGTATTG CAACGGGACT CAGCTTCGGC GAATAAGGGA GTCTACCCTA
b-ugd19   TTGCCAATAT CCTGGTATTG CAACGGGACT CGGCTTTGGC GAATAAGGGA GTCTAACCTA
 b-ugd7   TTGCCAATAT CCTGGTATTG CAACGGGACT CAGCTTTGGC GAATAACGGA CTCTACCCTA
 b-ugd2   TTGCCAATAT CCTGGTATTG CAACGGGACT CAGCTTTCGC GAATAAGGGC GTCTACCCTA
b-ugd21   TTGCCAATAT CCTGGTATTG CAACGGGACT CAGCTTTGGC GAATAAGGGA GAATACCCTA
b-ugd13   TTGCCAATAT CCTGGTATTG CAACGGGACT CGGCTTTGGC GAATAACGGC GTCTAACCTA
 b-ugd3   TTGCCAATAT CCTGGTATTG CAACGGGACT CGGCTTTGGC GAATAACGGC GTCTAACCTA
b-ugd15   TTGCCAATAT CCTGGTATTG CAACGGGACT CGGCTTTGGC GAATAACGGC GTCTAACCTA
b-ugd10   TTGCCAATAT CCTGGTATTG CAACGGGACT CAGCTTCACC GAATAACGGC GTCTAACCTA
b-ugd29   TTGCCAATAT CCTGGTATTG CAACGGGACT CGGCTTCACC GAATAACGGC GTCTAACCTA
a-ugd16   TTGCCAATAT CCTGGTATTG CAACGATACT CGGCTTTTGC GACTGACGGA CCAATCATTC
a-ugd18   TTGCCAATAT CCTGGTATTG CAACGATACT CGGCTTTTGC GACTGACGGA CCAATCATTC
a-ugd24   TTGCCAATAT CCTGGTATTG CAACGATACT CGGCTTTTGC GACTGACGGA CCAATCATTC
 a-ugd4   TTGCCAATAT CCTGGTATTG CAACGATACT CGGCTTTGGC GACTGACGGA CCAATCATTC
c-ugd23   TTGCCAATAT CCTGGTATTG CAACGATACT CGGCTTCTCC GACTGACGCA GCAAACATTC
a-ugd12   TTGCCAATAT CCTGGTATTG CAACGATACT CGGCTTTGGC GACTGACGGA CCAATCATTC
   Cons   TTGCCAA-AT CCTGGTATTG CAACG--ACT C-GCTT---C GA-T---G-- --------T-
```

```
          181                                                        240
 b-ugd1   ACTGTTTCTT TGCTCACCTG CAATTTTACT TGTATGACAG CATCTGGGCC TACACACAGC
b-ugd26   ACTGTTTCTT TCCTCACCTG CAATTTTACT TGTATGACAG CATCTGGGCC TACACACAGC
b-ugd19   ACTGTTGCTT CGCTCACCGG CAATTTTACT TGTATGACAG CATCTGGGCC TACACACAGC
 b-ugd7   ACTGTTTCTT CGCTCACCGG CAATTTTACT TGTATGACAG CATCTGGGCC TACACACAGC
 b-ugd2   ATTGTTTCTT CGCTCACCGG CAATTTTACT TGTATGACAG CATCTGGGCC TACACACAGC
b-ugd21   ACTGTTTCTT CGCTCACCGG CAATTTTACT TGTATGACAG CATCTGGGCC TACATACAGC
b-ugd13   ACTGTTTTGT CGGTGACCTG CAATTTTACT TGTATGACAA CATCTGGGCC TACACACAGC
 b-ugd3   ACTGTTTTGT CGGTGACCTG CAATTTTACT TGTATGACAA CATCTGGGCC TACACACAGC
b-ugd15   ACTGTTTTGT CGGTGACCTG CAATTTTACT TGTATGACAA CATCTGGGCC TACACACAGC
b-ugd10   ACTGTTTCTT CGCTCACCTG CAATTTTACT TGTATGACAA CATCTGGGCC TACACACAGC
b-ugd29   ACTGTTTCTT CGCTCACCTG CAATTTTACT TGTATGACAA CATCTGGGCC TACACACAGC
a-ugd16   ACTGTTGCCA CCATTACCTG CAATTTTACT TGTGTGGAAC AATCTGGGCA TCGACAGAGC
a-ugd18   ACTGTTGCCA CCATTACCTG CAATTTTACT TGTGTGGAAC AATCTGGGCA TCGACAGAGC
a-ugd24   ACTGTTGCCA CCATTACCTG CAATTTTACT TGTGTGGAAC AATCTGGGCA TCGACAGAGC
 a-ugd4   ACTGTTGCCA CCATTACCTG CAATTTTACT TGTGTGGAAC AATCTGGGCA TCGACAGAGC
c-ugd23   ACTTTTGCCA ACCTTACCTG CAATTTTACT TGTGTGGGAC AATCTGGGTA TCGACACAGC
a-ugd12   ACTGTTGCCA ACATTACCTG CAATTTTACT TGTGTGGAAC AATCTGGGCA TCGACAGAGC
   Cons   A-T-TT---- ---T-ACC-G CAATTTTACT TGT-TG--A- -ATCTGGG-- T--A-A-AGC

          241                                                        300
 b-ugd1   ATTTGGATTG AATGGTATAC AACACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
b-ugd26   ATTTGGATTG AATGGTATAC AACACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
b-ugd19   ATTTGGATTG AATGGTATAC AACACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
 b-ugd7   ATTTGGATTG AATGGTATAC AACACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
 b-ugd2   ATTTGGATTG AATGGCATAC AACACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
b-ugd21   ATTTGGATTG AATGGTATCC AACACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
b-ugd13   ATTTGGATTG AATGGTATTC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
 b-ugd3   ATTTGGATTG AATGGTATTC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
b-ugd15   ATTTGGATTG AATGGTATTC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
b-ugd10   ATTTGGATTC AATGGTATAC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
b-ugd29   ATTTGGATTG AATGGTATAC ACAACCTGTC TTACAAACCT TGTGTACACA GCCATCAAAC
a-ugd16   ATTTGGATTA CATGGAATGC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
a-ugd18   ATTTGGATTA CATGGAATGC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
a-ugd24   ATTTGGATTA CATGGAATGC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
 a-ugd4   ATTTGGATTA CATGGAATGC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
c-ugd23   GTTTGGATTA CATGGCATGC ACAACCTGTC TTACAAACCT TGTGTGCGCA ACCATCAAAC
a-ugd12   ATTTGGATTA CATGGAATGC ACAACCTGTC TTACAAACCT TGTGTGCACA GCCATCAAAC
   Cons   -TTTGGATT- -ATGG-AT-C A--ACCTGTC TTACAAACCT TGTGT-C-CA -CCATCAAAC

          301                                                        360
 b-ugd1   ACAGTGACTT GTGGTCAGCG TGTTACTTTG TATTGTCATA CCTCTTCAAA TAATGGTACC
b-ugd26   ACAGTGACTT GTGGTCAGCG TGTTACTTTG TATTGTTATA CCTCTTCAAA TAATGGTACC
b-ugd19   ACAGTGACTT GTGGTCAGCG TGTTACTTTG TATTGTCATA CCTCTTCAAA TAATGGTACC
 b-ugd7   ACAGTGACTT GTGGTCAGCG TGTTACTTTG TATTGTCATA CCTCTTCAAA TAATGGTACC
 b-ugd2   ACAGTGACTT GTGGTCAGCG TGTTACTTTG TATTGTCATA CCTCTTCAAA TAATGGTACC
b-ugd21   ACAGTGACTT GTGGTCAGCG TGTTACTTTG TATTGTCATA CCTCTTCAAA TAATGGTACC
b-ugd13   ACAGTGACTT GCGGTCAGCG TGTTACTTTG TATTGTCATA CCTCTTCAGA TAATGTTACC
 b-ugd3   ACAGTGACTT GCGGTCAGCG TGTTACTTTG TATTGTCATA CCTCTTCAGA TAATGTTACC
b-ugd15   ACAGTGACTT GCGGTCAGCG TGTTACTTTG TATTGTCATA CCTCTTCAGA TAATGTTACC
b-ugd10   ACAGTGTCTT GTGGTCAGCC TGTTACTTTG TATTGTGATA CCTCTTCAAA TAATGTTACC
b-ugd29   ACAGTGTCTT GTGGTCAGCC TGTTACTTTG TATTGTGATA CCTCTTCAAA TAATGTTACC
a-ugd16   ACAGTCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
a-ugd18   ACAGTCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
a-ugd24   ACAGTCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
 a-ugd4   ACAATCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
c-ugd23   ACAGTCACTT GTGGTCAGCA TGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
a-ugd12   ACAGTCGCTT GTGGTCAGCA GGTTACTTTG TATTGTTCTA CCTCTGGAAA TAATGTTACC
   Cons   ACA-T--CTT G-GGTCAGC- -GTTACTTTG TATTGT--TA CCTCT--A-A TAATG-TACC
```

```
          361                                                        420
 b-ugd1   ATTTGGCGTC TACAAAACGG ACACAATCAA ACCGTGTCAC AAACTAAATA CTATAATTTT
b-ugd26   ATTTGGCGTC TACAAAACGG ACACAATCAA ACCGTGTCAC AAACTAAATA CTATAATTTT
b-ugd19   ATTTGGCGTC TACAAAAGGG ACACAATCAA ACCGTGTCAC AAACTAAATA CTATAATTTT
 b-ugd7   ATTTGGCGTC TACAAAACGG ACACAATCAA ACCGTGTCAC AAACTAAATA CTATAATTTT
 b-ugd2   ATTTGGCGTC TACAAAACGG ACACAATCAA ACCGTGTCAC AAACTAAATA CTATAATTTT
b-ugd21   ATTTGGCGTC TACAAAACGG ACACAATCAA ACCGTGTCAC AAACTAAATA CTATAATTTT
b-ugd13   ATTTGGCGTC TACAAAACGG ACACAATCAA ACCGTGTCAC AAACTAAATA CTATAATTTT
 b-ugd3   ATTTGGCGTC TACAAAACGG ACACAATCAA ACCGTGTCAC AAACTAAATA CTATAATTTT
b-ugd15   ATTTGGCGTC TACAAAACGG ACACAATCAA ACCGTGTCAC AAACTAAATA CTATAATTTT
b-ugd10   ATTTGGCGTC TACAAAACGG ACACAATCAA ACCGTGTCAC AAACTAAATA CTATAATTTT
b-ugd29   ATTTGGCGTC TACAAACCGG ACACAATCAA ACCGTGTCAC AAACTAAATA CTATAATTTT
a-ugd16   GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
a-ugd18   GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
a-ugd24   GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
 a-ugd4   GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
c-ugd23   GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
a-ugd12   GTTTGGCATC TACCAAACGG ACGAAATGAA ACCGTGTCAC AAACTAAATA CTATAATTTT
   Cons   -TTTGGC-TC TAC-AA--GG AC--AAT-AA ACCGTGTCAC AAACTAAATA CTATAATTTT

          421                                                        480
 b-ugd1   ACGATGATGA ACCAAACTCA GGGGTGTTAT CTTTGTTCTG ACGGGCTGTC GTCTCGCCTG
b-ugd26   ACGATGATGA ACCAAACTCA GGGGTGTTAT CTTTGTTCTG ACGGGCTGTC GTCTCGCCTG
b-ugd19   ACGATGATGA ACCAAACTGA GGGGTGTTAT CTTTGTTCTG ACGGGCTGTC GTCTCGCCTG
 b-ugd7   ACGATGATGA ACCAAACTCA GGGGTGTTAT TTTTGTTTTG ACGGGCTGTC GTCTCGCCTG
 b-ugd2   ACGATGATGA ACCAAACTCA GGGGTGTTAT ATTTGTTGTG ACGGGCTGTC GTCTCACCTG
b-ugd21   ACGATGATGA ACCAAACTCA GGGGTGTTAT ATTTGTTGTG ACGGGCTGTC GTCTCGCCTG
b-ugd13   ACGATGATGG ACCAAACTCA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
 b-ugd3   ACGATGATGG ACCAAACTCA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
b-ugd15   ACGATGATGA ACCAAACTCA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
b-ugd10   ACGATGATGA ACCAAACTCA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
b-ugd29   ACGATGATGA ACCAAACTCA GGGGTGTTAT GCTTGTTCTA ACGGGCTGTC GTCTCGCCTG
a-ugd16   ACGCTGATGA ACCAAACTGA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
a-ugd18   ACGCTGATGA ACCAAACTGA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
a-ugd24   ACGCTGATGA ACCAAACTGA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
 a-ugd4   ACGCTGATGA ACCAAACTGA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
c-ugd23   ACGCTGATGA ACCAAACTGA GGGGTGTTAT GGTTGTTCTA ACGGCCTGTC GTCTCGCCTG
a-ugd12   ACGCTGATGA ACCAAACTGA GGGGTGTTAT ACTTGTTCTA ACGGGCTGTC GTCTCGCCTG
   Cons   ACG-TGATG- ACCAAACT-A GGGGTGTTAT --TTGTT-T- ACGG-CTGTC GTCTC-CCTG

          481                                                        540
 b-ugd1   TCAAATCGTA TATGTTTTTG GGCGTCTTGT ATCAATATAA CTCCAGAAAC TCCTACTGTC
b-ugd26   TCAAATCGTA TATGTTTTTG GGCGTCTTGT ATCAATATAA CTCCAGAAAC TCCTACTGTC
b-ugd19   TCAAATCGTA TATGTTTTTG GGCGTCTTGT ATCAATATAA CTCCAGAAAC TCCTACTGTC
 b-ugd7   TCAAATCGTA TATGTTTTTG GGCGTCTTGT ATCAATATAA CTCCAGAAAC TCCTACTGTC
 b-ugd2   TCAAATCGTA TATGTTTTTG GGCGTCTTGT ATCAATATAA CTCCAGAAAC TCCTACTGTC
b-ugd21   TCAAATCGTA TATGTTTTTG GGCGTCTTGT ATCAATATAA CTCCAGAAAC TCCTACTGTC
b-ugd13   TCAAATCGTA TATGTTTTTG GGCGTCTTGT ATTAATATAA CTCCAGAAAC TCCTACTGTC
 b-ugd3   TCAAATCGTA TATGTTTTTG GGCGTCTTGT ATTAATATAA CTCCAGAAAC TCCTACTGTC
b-ugd15   TCAAATCGTA TATGTTTTTG GGCGTCTTGT ATTAATATAA CTCCAGAAAC TCCTACTGTC
b-ugd10   TCAAATCGTA TATGTTTTTG GGCGTCTTGT ATCAATATAA CTCCAGAAAC TCCTACTGTC
b-ugd29   TCAAATCGTA TATGTTTTTG GGCGTCTTGT ATCAATGTAA CTCCAGAAAC TCCTACTGTC
a-ugd16   TCAAATCGTA TATGTTTTTG GGCGCGTTGT GCCAATATAA CTCCAGAAAC TCATACTGTA
a-ugd18   TCAAATCGTA TATGTTTTTG GGCGCGTTGT GCCAATATAA CTCCAGAAAC TCATACTGTA
a-ugd24   TCAAATCGTA TATGTTTTTG GGCGCGTTGT GCCAATATAA CTCCAGAAAC TCATACTGTA
 a-ugd4   TCAAATCGTA TATGTTTTTG GGCGCGTTGT GCCAATATAA CTCCAGAAAC TCATACTGTA
c-ugd23   TCAAATCGTA TATGTTTTTG GGCGCGTTGT GCCAATATAA CTCCAGAAAC TCATACTGTA
a-ugd12   TCAAATCGTA TATGTTTTTG GGCGCGTTGT GCCAATATAA CTCCAGAAAC TCATACTGTA
   Cons   TCAAATCGTA TATGTTTTTG GGCG--TTGT ---AAT-TAA CTCCAGAAAC TC-TACTGT-
```

```
         541                                                        600
 b-ugd1  TCCGCCAGCA GTACTACGGC CTTTAAAATG TCAAGAACTA ATGGATTACT GAAAATAATC
b-ugd26  TCCGCCAGCA GTACTACGGC CTTTAAAATG TCAAGAACTA ATGGATTACT GAAAATAATC
b-ugd19  TCCGCCAGCA GTACTACGGC CTTTAAAATG TCAAGAACTA ATGGATTACT GAAAATAATC
 b-ugd7  TCCGCCAGCA GTACTACGGC CTTTAAAATG TCAAGAACTA ATGGATTACT GAAAATAATC
 b-ugd2  TCCGCCAGCA GTACTACGGC CATTAAAATG TCAAGAACTA ATGGATTACT GAAAATAATC
b-ugd21  TCCGCCAGCA GTACTACGGC CATTAAAATG TCAAGAACTA ATGGATTACT GAAAATAATC
b-ugd13  TCCGCCAGCA GTACTACGGC CTTTAAAACA TTAACAACTA ATGGATTACT GAAAATAATC
 b-ugd3  TCCGCCAGCA GTACTACGGC CTTTAAAACA TTAACAACTA ATGGATTACT GAAAATAATC
b-ugd15  TCCGCCAGCA GTACTACGGC CTTTAAAACA TTAACAACTA ATGGATTACT GAAAATAATC
b-ugd10  TCCGCCAGCA GTACTATGGC CGTTAAACTA TTAAGAACTA ATGGATTACT GAAAATAATC
b-ugd29  TCCGCCAGCA GTCCTATGGC CGTTAAAGTA TTAAGAACTA ATGGATTACT GAAAATAATC
a-ugd16  TCTGTCAGCA GTACTACAGG CTTTAGAACA GTCAGTACTA ATAGCTTAGT GAACATAATC
a-ugd18  TCTGTCAGCA GTACTACAGG CTTTAGAACA GTCAGTACTA ATAGCTTAGT GAACATAATC
a-ugd24  TCTGTCAGCA GTACTACAGG CTTTAGAACA GTCAGTACTA ATAGCTTAGT GAACATAATC
 a-ugd4  TCTGTCAGCA GTACTACAGG CTTTAGAAAT TTAAGTACTC ATAGCTTAGT GAACATAATC
c-ugd23  TCTGTCAGCA GTACTACAGG CTTTGGAACA TTGAGTACTC ATAGCTTAGT GAACAGAATC
a-ugd12  TCTGTCAGCA GTACT..... .......... .......... .......... ....ATAATC
   Cons  TC-G-CAGCA GT-CT----- ---------- ---------- ---------- ----A-AATC

         601                                                        660
 b-ugd1  CCTGCAACCA CACATGCTGC AGTTGCAGTG GAAGAAGTAA AATCTACAAA TCCACATATT
b-ugd26  CCTGCAACCA CACATGCTGC AGTTGCAGTG GAAGAAGTAA AATCTACAAA TCCACATATT
b-ugd19  CCTGCAACCA CACATGCTGC AGTTGCAGTG GAAGAAGTAA AATCTACAAA TCCACATATT
 b-ugd7  CCTGCAACCA CACATGCTGC AGTTGCAGTG GAAGAAGTAA AATCTACAAA TCCACATATT
 b-ugd2  CCTGCAACCA CACATGCTGC AGTTGCAGTG GAAGAAGTAA AATCTACAAA TCCACATATT
b-ugd21  CCTGCAACCA CACATGCTGC AGTTGCAGTG GAAGAAGTAA AATCTACAAA TCCACATATT
b-ugd13  CCTGCAACCA CACATGCTGC AGTTGCAGTG GAAGAAGTAA AATCTACAAA TCCACATATT
 b-ugd3  CCTGCAACCA CACATGCTGC AGTTGCAGTG GAAGAAGTAA AATCTACAAA TCCACATATT
b-ugd15  CCTGCAACCA CACATGCTGC AGTTGCAGTG GAAGAAGTAA AATCTACAAA TCCACATATT
b-ugd10  CCTGCAACCA CACATGCTGC AGTTGCAGTG GAAGAAGTAA AATCTACAAA TACACATATT
b-ugd29  CCTGCAACCA CACATGCTGC AGTTGCAGTG GAAGAAGTAA AATCTACAAA TACACATATT
a-ugd16  CATGCAACCA ACCATGATGT AGTTGTAGTG AAAGAAGCAA AATCTACAAA TCCACATATT
a-ugd18  CATGCAACCA ACCATGATGT AGTTGTAGTG AAAGAAGCAA AATCTACAAA TCCACATATT
a-ugd24  CATGCAACCA CACATGATGT AGTTGTAGTG AAAGAAGCAA AATCTACAAA TCCACATATT
 a-ugd4  CATGCAACCA CACATCATGT AGTTGTAGTG AAAGAAGCAA AATCTACAAA TCCACATATT
c-ugd23  CATGCAACCA CACATGATGT AGTTGTAGTG AAAGAAGCAA AATCTACAAA TCCACATATT
a-ugd12  CATGCAACCA CACATGATGT AGTTGTAGTG AAAGAAGCAA AATCTACAAA TCCACATATT
   Cons  C-TGCAACCA --CAT--TG- AGTTG-AGTG -AAGAAG-AA AATCTACAAA T-CACATATT

         661                                                        720
 b-ugd1  CAAGTGCCTT TTCTTGTATT TATGACGCTC GTAGCTCTGA TAGGAACCAT GTGTGGTATC
b-ugd26  CAAGTGCCTT TTCTTGTATT TATGACGCTC GTAGCTCTGA TAGGAACCAT GTGTGGTATC
b-ugd19  CAAGTGCCTT TTCTTGTATT TATGACGCTC GTAGCTCTGA TAGGAACCAT GTGTGGTATC
 b-ugd7  CAAGTGCCTT TTCTTGTATT TATGACGCTC GTAGCTCTGA TAGGAACCAT GTGTGGTATC
 b-ugd2  CAAGTGCCTT TTCTTGTATT TATGACGCTC GTAGCTCTGA TAGGAACCAT GTGTGGTATC
b-ugd21  CAAGTGCCTT TTCTTGTATT TATGACGCTC GTAGCTCTGA TAGGAACCAT GTGTGGTATC
b-ugd13  CAAGTGCCTT TTCTTGTATT TATGACGCTC GTAGCTCTGA TAGGAACCAT GTGTGGTATC
 b-ugd3  CAAGTGCCTT TTCTTGTATT TATGACGCTC GTAGCTCTGA TAGGAACCAT GTGTGGTATC
b-ugd15  CAAGTGCCTT TTCTTGTATT TATGACGCTC GTAGCTCTGA TAGGAACCAT GTGTGGTATC
b-ugd10  CAAGTGCCTT TTCTTGTATT TATGACGCTC GTAGCTCTGA TAGGAACCAT GTGTGGTATC
b-ugd29  CAAGTGCCTT TTCTTGTATT TATGACGCTC GTAGCTCTGA TAGGAACCAT GTGTGGTATC
a-ugd16  GAAGTGCCTT TTCTTGTATT TATGACACTC GTAGCTCTGA TAGGAACCAT GTGTGGTATC
a-ugd18  GAAGTGCCTT TTCTTGTATT TATGACACTC GTAGCTCTGA TAGGAACCAT GTGTGGTATC
a-ugd24  GAAGTGCCTT TTCTTGTATT TATGACACTC GTAGCTCTGA TAGGAACCAT GTGTGGTATC
 a-ugd4  GAAGTGCCTT TTCTTGTATT TATGACACTC GTAGCTCTGA TAGGAACCAT GTGTGGTATC
c-ugd23  GAAGTGCCTT TTCTTGTATT TATGACACTC GTAGCTCTGA TAGGAACCAT GTGTGGTATC
a-ugd12  GAAGTGCCTT TTCTTGTATT TATGACACTC GTAGCTCTGA TAGGAACCAT GTGTGGTATC
   Cons  -AAGTGCCTT TTCTTGTATT TATGAC-CTC GTAGCTCTGA TAGGAACCAT GTGTGGTATC
```

```
          721                                                        780
 b-ugd1   TTAGGAACTC TTATCTTTAC CCATTGTCAA AAAAAAAGTG ACTCAAGCAA AACAGGGCAA
b-ugd26   TTAGGAACTC TTATCTTTAC CCATTGTCAA AAAAAAAGTG ACTCAAGCAA AACAGGGCAA
b-ugd19   TTAGGAACTC TTATCTTTAC CCATTGTCAA AAAAAAAGTG ACTCAAGCAA AACAGGGCAA
 b-ugd7   TTAGGAACTC TTATCTTTAC CCATTGTCAA AAAAAAAGTG ACTCAAGCAA AACAGGGCAA
 b-ugd2   TTAGGAACTC TTATCTTTAC CCATTGTCAA AAAAAAAGTG ACTCAAGCAA AACAGGGCAA
b-ugd21   TTAGGAACTC TTATCTTTAC CCATTGTCAA AAAAAAAGTG ACTCAAGCAA AACAGGGCAA
b-ugd13   TTAGGAACTG TTATCTTTAC CCATTGTCAA AAAAAAAGTG ACTCAAGCAA AACAGGGCAA
 b-ugd3   TTAGGAACTG TTATCTTTAC CCATTGTCAA AAAAAAAGTG ACTCAAGCAA AACAGGGCAA
b-ugd15   TTAGGAACTG TTATCTTTAC CCATTGTCAA AAAAAAAGTG ACTCAAGCAA AACAGGGCAA
b-ugd10   TTAGGAACTC TTATCTTTAC CCATTGTCAA AAAAAAAGTG ACTCAAGCAA AACAGGGCAA
b-ugd29   TTAGGAACTC TTATCTTTAC CCATTGTCAA AAAAAAAGTG ACTCAAGCAA AACAGGGCAA
a-ugd16   TTAGGAACTA TTATCTTTGC CCATTGTCAA AAACAACGTG ACTCAAACAA AACAGTGCCA
a-ugd18   TTAGGAACTA TTATCTTTGC CCATTGTCAA AAACAACGTG ACTCAAACAA AACAGTGCCA
a-ugd24   TTAGGAACTA TTATCTTTGC CCATTGTCAA AAACAACGTG ACTCAAACAA AACAGTGCCA
 a-ugd4   TTAGGAACTA TTATCTTTGC CCATTGTCAA AAACAACGTG ACTCAAACAA AACAGTGCCA
c-ugd23   TTAGGAACTA TTATCTTTGC CCATTGTCAA AAACAAAGTG ACTCAAACAA AACAGTGCCA
a-ugd12   TTAGGAACTA TTATCTTTGC CCATTGTCAA AAACAACGTG ACTCAAACAA AACAGTGCCA
   Cons   TTAGGAACT- TTATCTTT-C CCATTGTCAA AAA-AA-GTG ACTCAA-CAA AACAG-GC-A

          781                                                        840
 b-ugd1   CAACAATTGC GGGATTATTA TTCCCTAGAC TATTTTCACA CGGAAGAGTA TACGCAACCA
b-ugd26   CAACAATTGC GGGATTATTA TTCCCTAGAC TATTTTCACA CGGAAGAGTA TACGCAACCA
b-ugd19   CAACAATTGC GGGATTATTA TTCCCTAGAC TATTTTCACA CGGAAGAGTA TACGCAACCA
 b-ugd7   CAACAATTGC GGGATTATTA TTCCCTAGAC TATTTTCACA CGGAAGAGTA TACGCAACCA
 b-ugd2   CAACAATTGC GGGATTATTA TTCCCTAGAC TATTTTCACA CGGAAGAGTA TACGCAACCA
b-ugd21   CAACAATTGC GGGATTATTA TTCCCTAGAC TATTTTCACA CGGAAGAGTA TACGCAACCA
b-ugd13   CAACAATTGC GGGGTTATTA TTCCCTAGAC TATTTTCACA CGGAAGAGTA TACACAACCA
 b-ugd3   CAACAATTGC GGGGTTATTA TTCCCTAGAC TATTTTCACA CGGAAGAGTA TACACAACCA
b-ugd15   CAACAATTGC GGGATTATTA TTCCCTAGAC TATTTTCACA CGGAAGAGTA TACACAACCA
b-ugd10   CAACAATTGC GGGATTATTA TTCCCTAGAC TATTTTCACA CGGAAGACTA TACGCAACCA
b-ugd29   CAACAATTGC GGGATTATTA TTCCCTAGAC TATTTTCACA CGGAAGACTA TACGCAACCA
a-ugd16   CAACAATTGC AGGATTATTA TTCCCTACAC GATTTGTGCA CGGAAGACTA TACGCAACCA
a-ugd18   CAACAATTGC AGGATTATTA TTCCCTACAC GATTTGTGCA CGGAAGACTA TACGCAACCA
a-ugd24   CAACAATTGC AGGATTATTA TTCCCTACAC GATTTGTGCA CGGAAGACTA TACGCAACCA
 a-ugd4   CAACAATTGC AGGATTATTA TTCCCTACAC GATTTGTGCA CGGAAGACTA TACGCAACCA
c-ugd23   CAACAATTGC GGGATTATTA TTCCCTACAC GATTTCTGCA CGGAAGACTA TACGCAACCA
a-ugd12   CAACAATTGC AGGATTATTA TTCCCTACAC GATTTGTGCA CGGAAGACTA TACGCAACCA
   Cons   CAACAATTGC -GG-TTATTA TTCCCTA-AC -ATTT---CA CGGAAGA-TA TAC-CAACCA

          841          855
 b-ugd1   GTGGATTGGT ACTGA
b-ugd26   GTGGATTGGT ACTGA
b-ugd19   GTGGATTGGT ACTGA
 b-ugd7   GTGGATTGGT ACTGA
 b-ugd2   GTGGATTGGT ACTGA
b-ugd21   GTGGATTGGT ACTGA
b-ugd13   GTGGATTGGT ACTGA
 b-ugd3   GTGGATTGGT ACTGA
b-ugd15   GTGGATTGGT ACTGA
b-ugd10   GTGGAGTGGT ACTGA
b-ugd29   GTGGAGTGGT ACTGA
a-ugd16   GTGGATTGGT ACTGA
a-ugd18   GTGGATTGGT ACTGA
a-ugd24   GTGGATTGGT ACTGA
 a-ugd4   GTGGATTGGT ACTGA
c-ugd23   GTGGATTGGT ACTGA
a-ugd12   GTGGATTGGT ACTGA
   Cons   GTGGA-TGGT ACTGA
```

B

```
          1                                                          60
a-ugd16   VCSLAVCFQG LLSLYLQSSP NLCPGVISTP YTLTCLSDAS LPISWYCNDT RLLRLTDQSF
a-ugd18   VCSLAVCFQG LLSLYLQSSP NLCPGVISTP YTLTCLSDAS LPISWYCNDT RLLRLTDQSF
a-ugd24   VCSLAVCFQG LLSLYLQSSP NLCPGVISTP YTLTCLSDAS LPISWYCNDT RLLRLTDQSF
 a-ugd4   VCSLAVCFRG LLSLYLQSSP NLCPGVISTP YTLTCLSDAS LPISWYCNDT RLWRLTDQSF
a-ugd12   VCSLAVCFRG LLSLYLQSSP NLCPGVISTP YTLTCLSDAS LPISWYCNDT RLWRLTDQSF
c-ugd23   VCSLAVCFRG LLSLYVQSSP NLCPGVISTP YTLTCPSDAT LPISWYCNDT RLLRLTQQTF
 b-ugd2   LCSLVVWFPK LLSLHLPSFP HLCPGVHFTP YTLTCPSNRS LPISWYCNGT QLSRIRASTL
b-ugd21   LCSLVVWFPQ LLSLHLPSFP HLCPGVHFTS YTLTCPSNRS LPISWYCNGT QLWRIRENTL
 b-ugd1   LCCLVVWFPK LLSLHLPSFP HLCPGVLFTP YTLTCPSNRS LPTSWYCNGT QLRRIRGSNL
b-ugd26   LCSLVVWFPK LLSLHLPSFP HLCPGVLFTP YTLTCPSNRS LPISWYCNGT QLRRIRESTL
b-ugd19   LCSLVVWFPK LLSLHLPSFP HLCPGVLFTP YTLTCPSNRS LPISWYCNGT RLWRIRESNL
 b-ugd7   LCSLVVCFPK LLSLHLPSFP HLCPGVLFTP YTLTCPSNRS LPISWYCNGT QLWRITDSTL
b-ugd13   VCSLLVCFPK LLSLHLPSFP HLCPGVVSTP YTLTCPSNRS LPISWYCNGT RLWRITASNL
 b-ugd3   VCSLLVCFPK LLSLHLPSFP HLCPGVVSTP YTLTCPSNRS LPISWYCNGT RLWRITASNL
b-ugd15   VCSLLVCFPK LLSLHLPSFP HLCPGVVSTP YTLTCPSNRS LPISWYCNGT RLWRITASNL
b-ugd10   VCSLLVCFPK LLSLHLPSFP HLCPGVISTP YTLTCPSNRS LPISWYCNGT QLHRITASNL
b-ugd29   VSSLLVCFPK LLSLHLPSFP PLCPGVISTP YTLTCPSNRF LPISWYCNGT RLHRITASNL
   Cons   ---L-V-F-- LLSL---S-P -LCPGV--T- YTLTC-S--- LP-SWYCN-T -L-R------
                                                      <

          61                                                         120
a-ugd16   TVATITCNFT CVEQSGHRQS IWITWNAQPV LQTLCAQPSN TVTCGQHVTL YCSTSGNNVT
a-ugd18   TVATITCNFT CVEQSGHRQS IWITWNAQPV LQTLCAQPSN TVTCGQHVTL YCSTSGNNVT
a-ugd24   TVATITCNFT CVEQSGHRQS IWITWNAQPV LQTLCAQPSN TVTCGQHVTL YCSTSGNNVT
 a-ugd4   TVATITCNFT CVEQSGHRQS IWITWNAQPV LQTLCAQPSN TITCGQHVTL YCSTSGNNVT
a-ugd12   TVANITCNFT CVEQSGHRQS IWITWNAQPV LQTLCAQPSN TVACGQQVTL YCSTSGNNVT
c-ugd23   TFANLTCNFT CVGQSGYRHS VWITWHAQPV LQTLCAQPSN TVTCGQHVTL YCSTSGNNVT
 b-ugd2   IVSSLTGNFT CMTASGPTHS IWIEWHTTPV LQTLCAQPSN TVTCGQRVTL YCHTSSNNGT
b-ugd21   TVSSLTGNFT CMTASGPTYS IWIEWYPTPV LQTLCAQPSN TVTCGQRVTL YCHTSSNNGT
 b-ugd1   TVSLLTCNFT CMTASGPTHS IWIEWYTTPV LQTLCAQPSN TVTCGQRVTL YCHTSSNNGT
b-ugd26   TVSFLTCNFT CMTASGPTHS IWIEWYTTPV LQTLCAQPSN TVTCGQRVTL YCYTSSNNGT
b-ugd19   TVASLTGNFT CMTASGPTHS IWIEWYTTPV LQTLCAQPSN TVTCGQRVTL YCHTSSNNGT
 b-ugd7   TVSSLTGNFT CMTASGPTHS IWIEWYTTPV LQTLCAQPSN TVTCGQRVTL YCHTSSNNGT
b-ugd13   TVLSVTCNFT CMTTSGPTHS IWIEWYSQPV LQTLCAQPSN TVTCGQRVTL YCHTSSDNVT
 b-ugd3   TVLSVTCNFT CMTTSGPTHS IWIEWYSQPV LQTLCAQPSN TVTCGQRVTL YCHTSSDNVT
b-ugd15   TVLSVTCNFT CMTTSGPTHS IWIEWYSQPV LQTLCAQPSN TVTCGQRVTL YCHTSSDNVT
b-ugd10   TVSSLTCNFT CMTTSGPTHS IWIQWYTQPV LQTLCAQPSN TVSCGQPVTL YCDTSSNNVT
b-ugd29   TVSSLTCNFT CMTTSGPTHS IWIEWYTQPV LQTLCTQPSN TVSCGQPVTL YCDTSSNNVT
   Cons   -----T-NFT C---SG---S -WI-W---PV LQTLC-QPSN T--CGQ-VTL YC-TS--N-T
                 VR1                    >

          121                                                        180
a-ugd16   VWHLPNGRNE TVSQTKYYNF TLMNQTEGCY TCSNGLSSRL SNRICFWARC ANITPETHTV
a-ugd18   VWHLPNGRNE TVSQTKYYNF TLMNQTEGCY TCSNGLSSRL SNRICFWARC ANITPETHTV
a-ugd24   VWHLPNGRNE TVSQTKYYNF TLMNQTEGCY TCSNGLSSRL SNRICFWARC ANITPETHTV
 a-ugd4   VWHLPNGRNE TVSQTKYYNF TLMNQTEGCY TCSNGLSSRL SNRICFWARC ANITPETHTV
a-ugd12   VWHLPNGRNE TVSQTKYYNF TLMNQTEGCY TCSNGLSSRL SNRICFWARC ANITPETHTV
c-ugd23   VWHLPNGRNE TVSQTKYYNF TLMNQTEGCY GCSNGLSSRL SNRICFWARC ANITPETHTV
 b-ugd2   IWRLQNGHNQ TVSQTKYYNF TMMNQTQGCY ICCDGLSSHL SNRICFWASC INITPETPTV
b-ugd21   IWRLQNGHNQ TVSQTKYYNF TMMNQTQGCY ICCDGLSSRL SNRICFWASC INITPETPTV
 b-ugd1   IWRLQNGHNQ TVSQTKYYNF TMMNQTQGCY LCSDGLSSRL SNRICFWASC INITPETPTV
b-ugd26   IWRLQNGHNQ TVSQTKYYNF TMMNQTQGCY LCSDGLSSRL SNRICFWASC INITPETPTV
b-ugd19   IWRLQKGHNQ TVSQTKYYNF TMMNQTEGCY LCSDGLSSRL SNRICFWASC INITPETPTV
 b-ugd7   IWRLQNGHNQ TVSQTKYYNF TMMNQTQGCY FCFDGLSSRL SNRICFWASC INITPETPTV
b-ugd13   IWRLQNGHNQ TVSQTKYYNF TMMDQTQGCY TCSNGLSSRL SNRICFWASC INITPETPTV
 b-ugd3   IWRLQNGHNQ TVSQTKYYNF TMMDQTQGCY TCSNGLSSRL SNRICFWASC INITPETPTV
b-ugd15   IWRLQNGHNQ TVSQTKYYNF TMMNQTQGCY TCSNGLSSRL SNRICFWASC INITPETPTV
b-ugd10   IWRLQNGHNQ TVSQTKYYNF TMMNQTQGCY TCSNGLSSRL SNRICFWASC INITPETPTV
b-ugd29   IWRLQTGHNQ TVSQTKYYNF TMMNQTQGCY ACSNGLSSRL SNRICFWASC INVTPETPTV
   Cons   -W-L--G-N- TVSQTKYYNF T-M-QT-GCY -C--GLSS-L SNRICFWA-C -N-TPET-TV
```

```
          181                                                        240
a-ugd16   SVSSTTGFRT VSTNSLVNII HATNHDVVVV KEAKSTNPHI EVPFLVFMTL VALIGTMCGI
a-ugd18   SVSSTTGFRT VSTNSLVNII HATNHDVVVV KEAKSTNPHI EVPFLVFMTL VALIGTMCGI
a-ugd24   SVSSTTGFRT VSTNSLVNII HATTHDVVVV KEAKSTNPHI EVPFLVFMTL VALIGTMCGI
 a-ugd4   SVSSTTGFRN LSTHSLVNII HATTHHVVVV KEAKSTNPHI EVPFLVFMTL VALIGTMCGI
a-ugd12   SVSST..... ........II HATTHDVVVV KEAKSTNPHI EVPFLVFMTL VALIGTMCGI
c-ugd23   SVSSTTGFGT LSTHSLVNRI HATTHDVVVV KEAKSTNPHI EVPFLVFMTL VALIGTMCGI
 b-ugd2   SASSTTAIKM SRTNGLLKII PATTHAAVAV EEVKSTNPHI QVPFLVFMTL VALIGTMCGI
b-ugd21   SASSTTAIKM SRTNGLLKII PATTHAAVAV EEVKSTNPHI QVPFLVFMTL VALIGTMCGI
 b-ugd1   SASSTTAFKM SRTNGLLKII PATTHAAVAV EEVKSTNPHI QVPFLVFMTL VALIGTMCGI
b-ugd26   SASSTTAFKM SRTNGLLKII PATTHAAVAV EEVKSTNPHI QVPFLVFMTL VALIGTMCGI
b-ugd19   SASSTTAFKM SRTNGLLKII PATTHAAVAV EEVKSTNPHI QVPFLVFMTL VALIGTMCGI
 b-ugd7   SASSTTAFKM SRTNGLLKII PATTHAAVAV EEVKSTNPHI QVPFLVFMTL VALIGTMCGI
b-ugd13   SASSTTAFKT LTTNGLLKII PATTHAAVAV EEVKSTNPHI QVPFLVFMTL VALIGTMCGI
 b-ugd3   SASSTTAFKT LTTNGLLKII PATTHAAVAV EEVKSTNPHI QVPFLVFMTL VALIGTMCGI
b-ugd15   SASSTTAFKT LTTNGLLKII PATTHAAVAV EEVKSTNPHI QVPFLVFMTL VALIGTMCGI
b-ugd10   SASSTMAVKL LRTNGLLKII PATTHAAVAV EEVKSTNTHI QVPFLVFMTL VALIGTMCGI
b-ugd29   SASSPMAVKV LRTNGLLKII PATTHAAVAV EEVKSTNTHI QVPFLVFMTL VALIGTMCGI
   Cons   S-SS------ ---------I -AT-H--V-V -E-KSTN-HI -VPFLVFMTL VALIGTMCGI
             <              VR2            >          <          TM

          241                                                  284
a-ugd16   LGTIIFAHCQ KQRDSNKTVP QQLQDYYSLH DLCTEDYTQP VDWY
a-ugd18   LGTIIFAHCQ KQRDSNKTVP QQLQDYYSLH DLCTEDYTQP VDWY
a-ugd24   LGTIIFAHCQ KQRDSNKTVP QQLQDYYSLH DLCTEDYTQP VDWY
 a-ugd4   LGTIIFAHCQ KQRDSNKTVP QQLQDYYSLH DLCTEDYTQP VDWY
a-ugd12   LGTIIFAHCQ KQRDSNKTVP QQLQDYYSLH DLCTEDYTQP VDWY
c-ugd23   LGTIIFAHCQ KQSDSNKTVP QQLRDYYSLH DFCTEDYTQP VDWY
 b-ugd2   LGTLIFTHCQ KKSDSSKTGQ QQLRDYYSLD YFHTEEYTQP VDWY
b-ugd21   LGTLIFTHCQ KKSDSSKTGQ QQLRDYYSLD YFHTEEYTQP VDWY
 b-ugd1   LGTLIFTHCQ KKSDSSKTGQ QQLRDYYSLD YFHTEEYTQP VDWY
b-ugd26   LGTLIFTHCQ KKSDSSKTGQ QQLRDYYSLD YFHTEEYTQP VDWY
b-ugd19   LGTLIFTHCQ KKSDSSKTGQ QQLRDYYSLD YFHTEEYTQP VDWY
 b-ugd7   LGTLIFTHCQ KKSDSSKTGQ QQLRDYYSLD YFHTEEYTQP VDWY
b-ugd13   LGTVIFTHCQ KKSDSSKTGQ QQLRGYYSLD YFHTEEYTQP VDWY
 b-ugd3   LGTVIFTHCQ KKSDSSKTGQ QQLRGYYSLD YFHTEEYTQP VDWY
b-ugd15   LGTVIFTHCQ KKSDSSKTGQ QQLRDYYSLD YFHTEEYTQP VDWY
b-ugd10   LGTLIFTHCQ KKSDSSKTGQ QQLRDYYSLD YFHTEDYTQP VEWY
b-ugd29   LGTLIFTHCQ KKSDSSKTGQ QQLRDYYSLD YFHTEDYTQP VEWY
   Cons   LGT-IF-HCQ K--DS-KT-- QQL--YYSL- ---TE-YTQP V-WY
                              *  *        *  *
             ><                CYT                  >
```

Table 4.3. Divergence values between pairs of K1 DNA and protein sequences[a]

	Ugd16	Ugd18	Ugd24	Ugd12	Ugd4	Ugd23	Ugd1	Ugd26	Ugd19	Ugd7	Ugd2	Ugd21	Ugd13	Ugd3	Ugd15	Ugd10	Ugd29
Ugd16	.000	.000	.0025	.0086	.0074	.0355	.1601	.1489	.1491	.1472	.1569	.1541	.1455	.1455	.1425	.1416	.1526
Ugd18	.000	.000	.0025	.0086	.0074	.0355	.1601	.1489	.1491	.1472	.1569	.1541	.1455	.1455	.1425	.1416	.1526
Ugd24	.0013	.0013	.000	.0062	.0049	.0328	.1567	.1456	.1458	.1440	.1535	.1508	.1422	.1422	.1392	.1384	.1493
Ugd12	.0217	.0217	.0181	.000	.0062	.0342	.1584	.1472	.1490	.1471	.1585	.1539	.1453	.1453	.1423	.1415	.1524
Ugd4	.0181	.0181	.0145	.0181	.000	.0355	.1582	.1471	.1474	.1455	.1569	.1523	.1437	.1437	.1407	.1414	.1523
Ugd23	.0771	.0771	.0731	.0768	.0813	.000	.1616	.1504	.1552	.1566	.1630	.1617	.1563	.1563	.1533	.1444	.1523
Ugd1	.3499	.3499	.3445	.3483	.3493	.3513	.000	.0086	.0136	.0162	.0187	.0212	.0419	.0419	.0393	.0355	.0447
Ugd26	.3367	.3367	.3314	.3353	.3362	.3316	.0223	.000	.0124	.0124	.0162	.0174	.0419	.0419	.0393	.0342	.0434
Ugd19	.3324	.3324	.3271	.3274	.3283	.3329	.0379	.0337	.000	.0124	.0162	.0175	.0394	.0394	.0368	.0382	.0448
Ugd7	.3266	.3266	.3213	.3217	.3225	.3271	.0458	.0376	.0412	.000	.0137	.0149	.0380	.0380	.0354	.0342	.0434
Ugd2	.3629	.3629	.3574	.3607	.3615	.3579	.0493	.0412	.0485	.0448	.000	.0136	.0432	.0432	.0406	.0382	.0487
Ugd21	.3607	.3607	.3552	.3556	.3565	.3603	.0575	.0452	.0488	.0487	.0371	.000	.0459	.0459	.0433	.0435	.0542
Ugd13	.2969	.2969	.2918	.2921	.2928	.3093	.0976	.0970	.0890	.0814	.1049	.1094	.000	.000	.0025	.0302	.0380
Ugd3	.2969	.2969	.2918	.2921	.2928	.3093	.0976	.0970	.0890	.0814	.1049	.1094	.000	.000	.0025	.0302	.0380
Ugd15	.2864	.2864	.2813	.2816	.2823	.2985	.0892	.0887	.0807	.0732	.0966	.1010	.0073	.0073	.000	.0276	.0354
Ugd10	.2900	.2900	.2849	.2892	.2949	.2970	.0927	.0916	.1045	.0845	.1047	.1206	.0674	.0674	.0595	.000	.0124
Ugd29	.3246	.3246	.3193	.3236	.3296	.3209	.1220	.1206	.1215	.1134	.1350	.1514	.0915	.0915	.0833	.0366	.000

[a] Divergence values among the DNA sequences are indicated in the upper triangle of the table (i.e. to the right of the identity diagonal) while those among the protein sequences are indicated in the bottom triangle (i.e. to the left of the identity diagonal).

divergences (expressed as the percentage of substitutions per nucleotide or amino acid residue) within and between K1 subtypes were then calculated and are shown in Table 4.4. The sequences exhibit among themselves from 0% (Ugd16 vs. Ugd18 and Ugd3 vs. Ugd13) to 14.2% (Ugd2 vs. Ugd23) nucleotide divergence and from 0% (Ugd16 vs. Ugd18 and Ugd3 vs. Ugd13) to 30.5% (Ugd2 vs. Ugd16) amino acid divergence. Divergence within A5, BI, BII, and BIII groups is minimal in both the nucleotide (0-0.9%, 0.9-2.1%, 1.2% and 0-0.2%, respectively) and amino acid (0-2.2%, 2.1-5.6%, 3.9%, and 0-0.7%, respectively) sequences. Divergence between the B subdivisions is 2.7-5.1% in nucleotide sequence and 6.7-13.7% in amino acid sequence. A high divergence of up to 14.2% and 30.5% in nucleotide and amino acid sequences, respectively, is noted between subtype B and A5 and C (Ugd23) variants. Divergence between A5 and C variants is much less, up to 3.4% and 8.1% in nucleotide and amino acid sequences, respectively, indicating that Ugd23 is closely related to the A5 variants. Indeed, pairwise divergences in DNA sequences of the 52 strains indicates that A5 strains and Ugd23 are more widely diverged from other A strains (up to 5% and 6.3%, respectively) than they are from each other. Furthermore, Ugd23 DNA sequence differs from other C subtype sequences by up to 6%. DNA sequence divergence among the B variants for the 52-sequence set is the same as that in the 17-sequence set of Ugandan strains only.

Visual inspection of the predominantly B subtype alignment comprising K1 amino acid sequences of the 17 Ugandan samples (Fig. 4.7B) indicates that the VR1 and VR2 regions are present, located between aa 53-94 and aa 190-219, respectively. The VR2 in this alignment is shorter than that in the alignment comprising the 52 sequences (aa 149-230; Fig. 4.5B). Peak variability within VR1 and VR2 is between aa 60-70 and aa 190-203, respectively. The alignment also shows a high level of variability across the entire gene, except for the transmembrane region (aa 230-253) which is highly conserved. Although the

Table 4.4. Range of pairwise divergence values[a] within and between K1 subtypes

Subtype	A5	C	BI	BII	BIII
A5	0-0.9 0-2.2	3.2-3.4 7.7-8.1	12.7-14.0 28.1-30.5	12.3-13.4 26.7-29.1	12.4-12.9 25.3-26.3
C		ND[b]	13.2-14.2 28.1-29.8	12.7-13.4 26.7-28.1	13.5-13.7 25.6-26.3
BI			0.9-2.1 2.1-5.6	3.3-5.1 8.8-13.7	3.4-4.4 7.7-10.9
BII				1.2 3.9	2.7-3.7 6.7-9.5
BIII					0-0.2 0-0.7

[a] Values are expressed as the percentage of substitutions per nucleotide (top value in cell) or amino acid residue (bottom value).

[b] Not done, only one C isolate was identified.

short cytoplasmic domain (37 residues) is also variable, the predicted ITAM motif ($DX_{10}DYYSLX_7YTQP$) is highly conserved. Cysteine residues noted previously to be conserved (Zong et al., 1999) are also conserved in this alignment, except two (aa 12 and aa 72) that were replaced with W or G, respectively (Fig. 4.7B).

4.7 ASSOCIATION OF K1 SUBTYPES WITH DISEASE AND TRIBE

The 17 K1 subtypes were distributed by HIV status, clinical presentation and tribe of the patients. Table 4.5A shows the association between subtypes and HIV status. Four of the 5 A5 subtypes came from HIV-negative individuals, and one from a HIV-positive patient. The B subtype was nearly equally distributed between the two groups, while the single C variant came from an HIV-positive individual. All the three subtypes were associated with nodular KS lesions and involvement of the extremities. The clinically aggressive non-AIDS-associated case (Ugd12) (Table 4.2) was associated with an A5 variant.

The 12 tribes represented were categorized into five groups according to the region in Uganda where the tribes occur predominantly (the total number of patients in each category is shown in parentheses): Central Uganda (7)-Ganda; Eastern Uganda (3)-Nyori/Jopadhola/Samia; Rwanda (3)-Nyarwanda/Mufumbira; South-West Uganda (3)-Nyankole/Mutoro/Konjo/Nyoro; Northern Uganda (1)-Lugbara/Luo/Langi. These groups broadly correspond to ethnic background. The results, presented in Table 4.5B, show that while the B subtype is present in all tribal groups, three of the five A5 variants came from patients of the Ganda tribe (Central Uganda) and two from patients belonging to the South-West Uganda group. The single C variant came from a Mufumbira, a tribe ethnically related to that of the people in Rwanda.

Table 4.5.

A. K1 subtypes and HIV status

Subtype	A5	B	C	TOTAL
HIV status				
HIV-1 positive	1	6	1	8
HIV-1 negative	4	5	0	9
TOTAL	5	11	1	17

B. K1 subtypes and tribe

Subtype	A5	B	C	TOTAL
Tribal group [a]				
Central Uganda	3	4	0	7
Eastern Uganda	0	3	0	3
Rwanda	0	2	1	3
South-West Uganda	2	1	0	3
Northern Uganda	0	1	0	1
TOTAL	5	11	1	17

[a] The tribal groups include: Central Uganda-Ganda; Eastern Uganda-Nyori/Jopadhola/Samia; Rwanda-Nyarwanda/Mufumbira; South-West Uganda-Nyankole/Mutoro/Konjo/Nyoro; Northern Uganda-Lugbara/Luo/Langi.

4.8 ASSOCIATION OF K1 SUBTYPES WITH PLACE OF RESIDENCE

The K1 subtypes were plotted on a map of Uganda according to the place of residence of the patient, as shown in Fig. 4.8. The B subtype is distributed throughout the region where patients came from (i.e. the southern part of the country). The few A5 variants are also distributed throughout this region, except for the Eastern part. The C variant came from Jinja, in the Eastern part of the country. A number of strains (8, two A5 and six B) cluster around the Kampala area.

4.9 CONCLUSION AND DISCUSSION

4.9.1 K1 subtypes

Representatives of three of the five subtypes of K1 gene were identified in biopsy samples from 17 Ugandan KS patients: five A5 variants, 11 subtype B strains and a single unusual subtype C variant. Subtype B was found to be more distant from A5 and C than A5 and C are from each other. These results are consistent with those of previous studies (Cook et al., 1999; Meng et al., 1999; Zong et al., 1999) and strengthen the evidence that subtype B predominates in Uganda. In contrast to the situation in Central and West Africa (Lacoste et al., 2000a), the A5 variant does not appear to be as prevalent as the B subtype in Uganda.

4.9.2 Association of K1 subtypes with disease, tribe and place of residence

As in most previous studies, no obvious association was noted between subtype B and HIV status, clinical picture, tribe or place of residence of the patient. However, a hint of a geographical association was noted in that all three samples

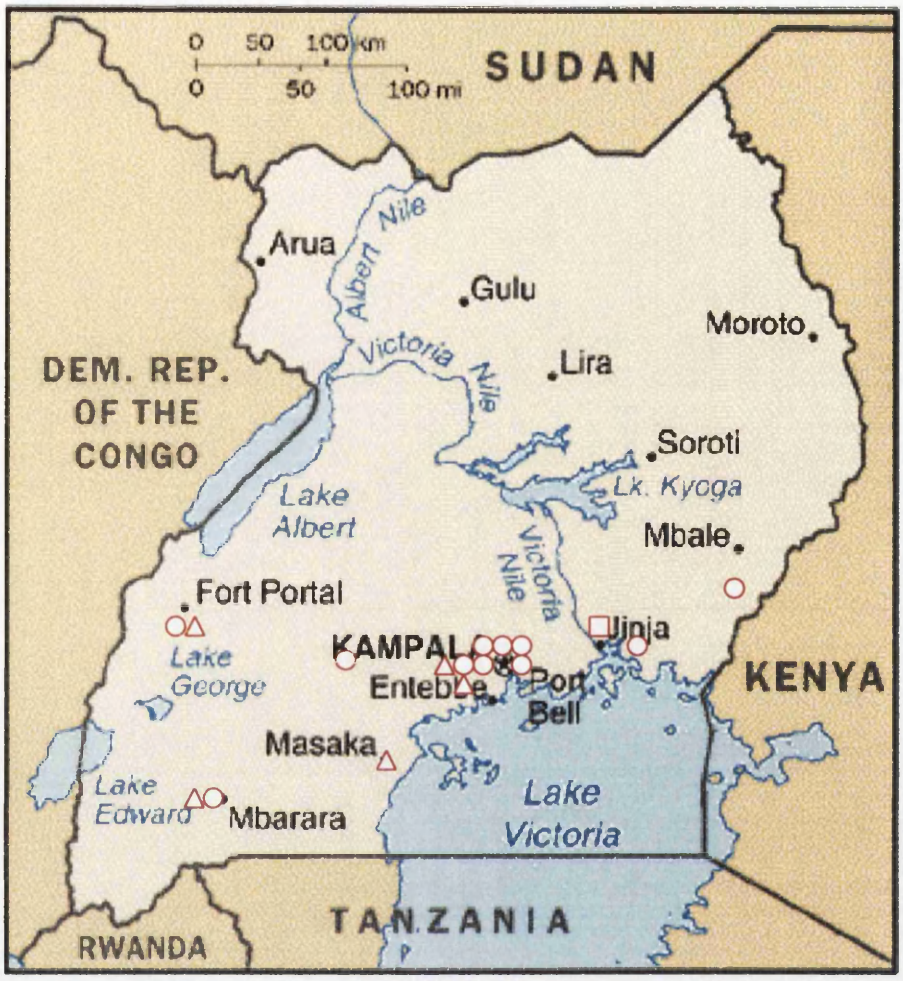


Fig. 4.8. Geographical distribution of K1 subtypes in Uganda.

The subtypes were plotted according to the place of residence of the KS patient. Triangles, circles and square represent A5, B and C subtypes, respectively.

from Mukono area (Ugd1, Ugd21 and Ugd26) belonged to one subtype B subdivision (BI; Table 4.2 and Fig. 4.6). It should be noted that the sample size in these analyses was too small (n=17) for meaningful statistical analysis to be performed. Thus, larger studies are required before definitive conclusions can be drawn.

The fact that no patients came from Northern Uganda was due to a war going on in that area during the time I collected samples, and not because KS is absent. A large proportion of patients (8) came from around Kampala because of easy access to UCI, and not because KS is more prevalent in this area.

CHAPTER 5
RESULTS AND DISCUSSION-III
CHARACTERIZATION OF THE K15 GENE

5.1 INTRODUCTION

The HHV-8 K15 gene occurs as two highly diverged alleles, P and M (Choi et al., 2000; Glenn et al., 1999; Poole et al., 1999), which appear to be essentially unlinked to K1 subtype (Poole et al., 1999). Divergence within each allele, however, appears to be minimal (Hayward, 1999). The M allele has been identified in samples from various parts of the world, including parts of Africa but not East Africa (Lacoste et al., 2000a,b; Meng et al., 2001; Poole et al., 1999). It appears to be highly prevalent in West and Central Africa (Lacoste et al., 2000a) where 10 of 19 samples contained this allele. Poole et al. (1999) detected the M allele in 18 of 63 HHV-8 strains from various parts of the world (mostly USA). It was most commonly associated with the K1 C subtype (especially from Taiwan); 13 of the 18 M-allele containing genomes had subtype C K1 genes. Poole et al. (1999) reported that there was virtually no sequence variation at all within the M allele of these strains, although the data were not published. In this study, eight strains from Eastern Africa (five from Zambia, two from Tanzania and one from Zaire) had the P allele. The K15 allele in the two strains from Uganda (ST1 and ST2) was not determined. Indeed, no study has examined the K15 allele(s) present in HHV-8 strains from Uganda. The objectives of this study were to identify the K15 allele(s) present in Ugandan strains and to evaluate divergence within them.

5.2 IDENTIFICATION OF K15 GENOTYPES IN UGANDAN SAMPLES

To identify the K15 alleles present in DNA samples from 30 KS patients Table 4.2), specific fragments of the K15 gene were amplified using either K15 M- (K15-3A/K15-4A) or P- (K15-3C/K15-4C) specific primers. Details of the primers used are given in Table 2.3. Twenty-seven samples gave products with P- but not M-specific primers, indicating that they contain the P allele. One sample (Ugd10) gave a product with M- but not P-specific primers, indicating that it contains the M allele. Two samples (Ugd5 and Ugd14) gave no product with either primer pair. Representative PCR products are shown in Fig. 5.1. All positive samples (except one) gave PCR products on the first attempt. The exception, Ugd30, required several attempts involving changes in PCR conditions. Eventually, a faint band was obtained when the annealing temperature was increased from 47.5 °C to 60 °C.

5.3 PCR AMPLIFICATION OF THE ENTIRE K15 GENE

The entire K15 gene of Ugandan samples representing the M (Ugd10; 2101 bp) and P (Ugd2, Ugd4, Ugd12, Ugd15, Ugd16, Ugd19, Ugd23 and Ugd29; up to 2086) alleles was amplified. The eight samples with the P allele represented the three K1 subtypes. Each PCR product also consisted of sequences (about 200 bp) flanking each side of the gene, with the downstream region extending to the beginning of ORF75. The sequence for the M allele was derived from two overlapping fragments (I and II) (Table 2.3). The PCR products are shown in Fig. 5.2A. Four primer sets were tried for amplification of the entire K15P gene, yielding varying levels of specificity. Primer set K15-6C/K5-7C (Table 2.3) gave the best results: fewer non-specific bands and stronger specific bands. The specific bands were gel purified and are shown in Fig. 5.2B. For both alleles, the PCR products were cloned before sequencing.

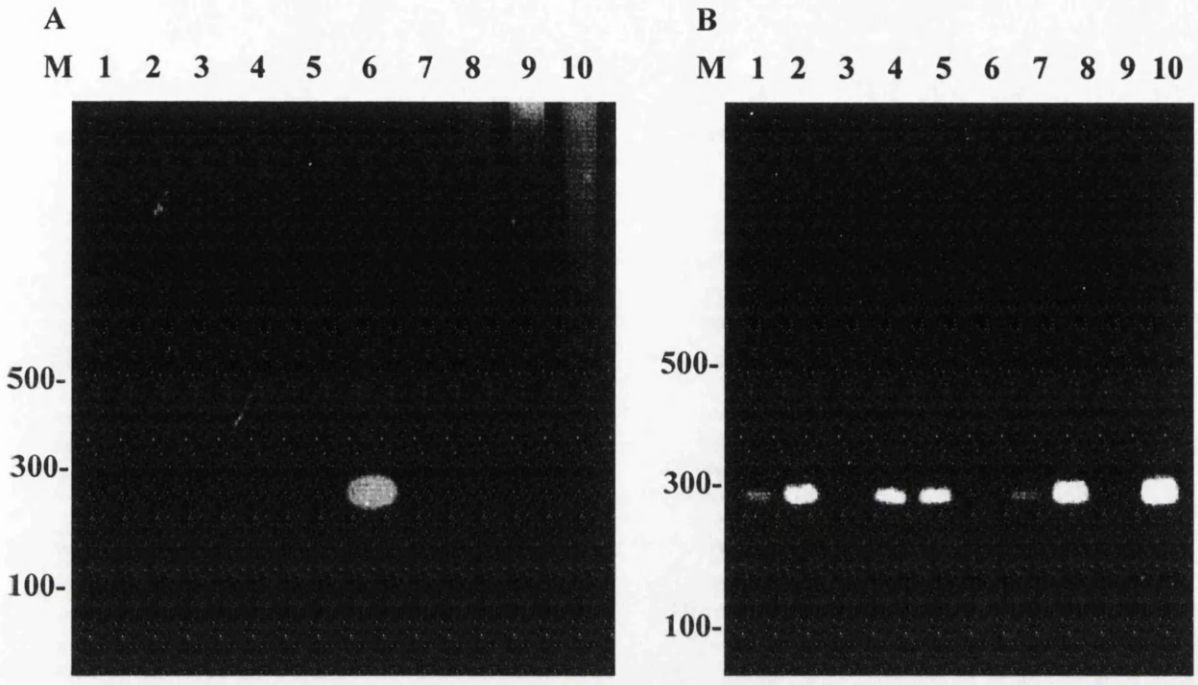


Fig. 5.1. K15 genotypes in KS tumour DNA samples.

EtBr-stained 1.8% (w/v) agarose gels showing PCR products of representative samples amplified with (A) K15 M- (K15-3A/4A; 298 bp) or (B) K15 P- (K15-3C/4C; 285 bp) specific primers. Lanes: 1, 2, 4, 5, 7, 8, and 10, representative samples; 6, Ugd10; 3, water control; 9, negative DNA control; M, 100 bp DNA ladder.

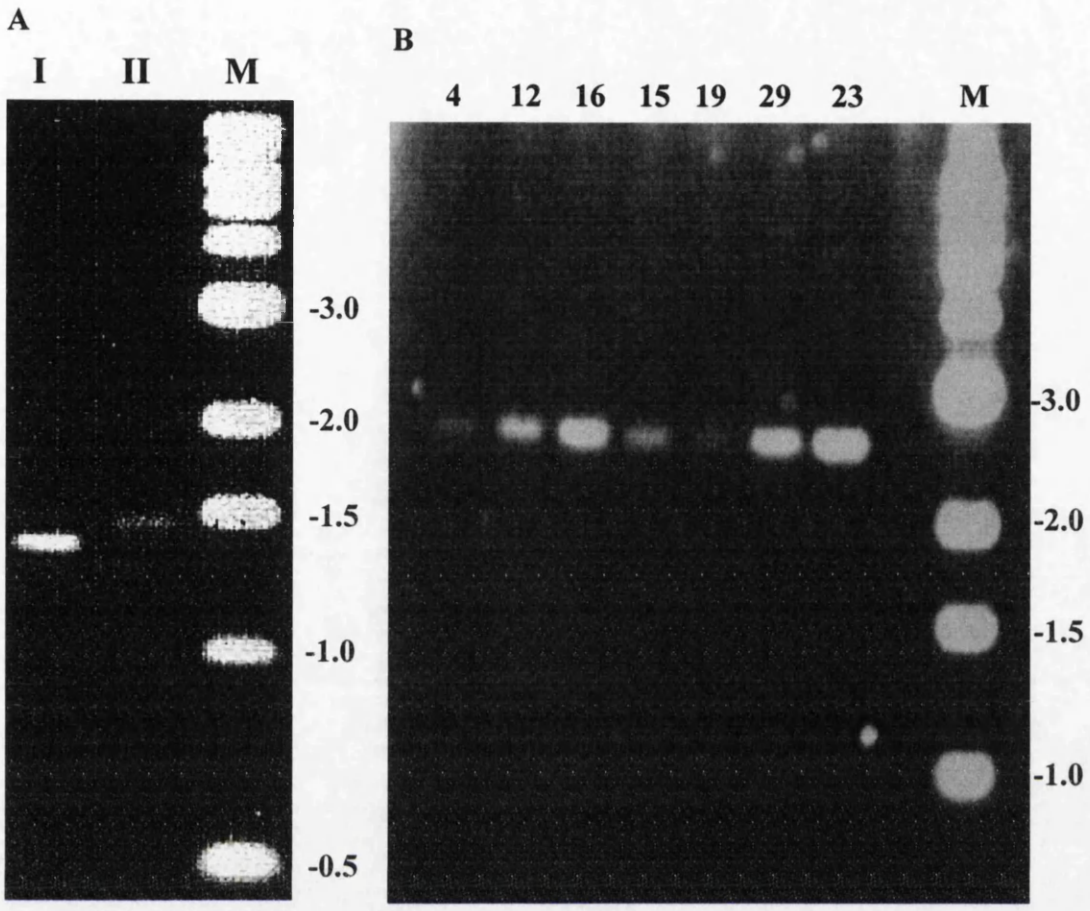


Fig. 5.2. PCR amplification of the entire K15 gene.

EtBr-stained 1% (w/v) agarose gels showing PCR products of (A) Ugd10 (M allele) and (B) representative samples containing the P allele. (A) The two overlapping fragments, I (1374 bp) and II (1483 bp) used to amplify the entire K15 gene of Ugd10. (B) The products (2494 bp) containing the full length K15 gene of the samples with the P allele. The numbers at the top of the lanes denote sample names without the Ugd-prefix, e.g. 4 is Ugd4.
M, 1 kbp DNA ladder.

5.4 DIVERGENCE WITHIN THE K15 GENE

5.4.1 Divergence within the K15 M allele

The sequence of Ugd10 was compared with that of BC-1, the only M allele sequence published. The DNA alignment (including K15 gene plus flanking sequences) is shown in Fig. 5.3, and the positions of nucleotide and amino acid changes (in the K15 gene alone) are shown in Fig. 5.4A.

The results show that the sequence of Ugd10 is closely related to that of BC-1. The two sequences are equal in length, except that BC-1 lacks a T residue within exon 1 (position 370 in Fig. 5.3; this was previously identified as an error; Glenn et al., 1999; Poole et al., 1999) and has an additional G residue in intron 5 (position 1416 in Figure 5.3). Both sequences exhibit major divergence from the P allele, commencing at the same position 21 bp downstream from the K15 stop codon (position 2331 in Fig. 5.3).

The K15 gene of Ugd10 differs from that of BC-1 at 25 positions (Fig. 5.4A). This is equivalent to a divergence of 1.2%. In coding regions, six substitutions are synonymous and nine are non-synonymous; ten substitutions are within introns. One substitution was identified in the upstream flanking region and none in the downstream region (Fig. 5.3).

5.4.2 Divergence within the K15 P allele

The K15 P sequences were analysed together with the corresponding regions of GK18 and BCBL-R (Table 2.1). The DNA sequence alignment of the K15 gene plus flanking regions is shown in Fig. 5.5. K15 P sequences exhibit major

Fig. 5.3. DNA alignment of K15 M allele sequences.

The alignment includes the entire K15 gene, and sequences (approx. 200 bp) flanking both sides, of Ugd10 and BC-1 oriented from left to right, i.e. opposite to the genomic orientation; nucleotides 1 and 2540 are equivalent to 136980 and 134442, respectively, in the BC-1 genome. The BC-1 sequence was modified before alignment by replacing the missing T residue (position 370 marked with asterisk). The start and stop codons are highlighted in bold-blue. Sequences in exons are underlined with a dashed line enclosed within '<' and '>' signs. Positions of nucleotide changes are shown in bold-red; the deletion is represented by a dot. The position (nucleotide 2331) where the M and P alleles become distinct is highlighted in bold-pink.

```
        1                                                           60
Ugd10   CACCAGGATG CAGTGTTACA TTATCAGCTT TTTGCAGCTG TGCCATTTTA TAAGCCCTTC
BC-1    CACCAGGATG CAGTGTTACA TTATCAGCTT TTTGCAGCTG TGCCATTTTA TAAGCCCTTC
Cons    CACCAGGATG CAGTGTTACA TTATCAGCTT TTTGCAGCTG TGCCATTTTA TAAGCCCTTC

        61                                                          120
Ugd10   ATCCGGATTT TAGGTTTCCA AACAAAAATA GGATGTCTGG TTCTGTGCAA AACTAACTCA
BC-1    ATCCGGATTT TAGGTTTCCA AACAAAAATA GGATGTCTGG TTCTGTGCAA AACTAACTCA
Cons    ATCCGGATTT TAGGTTTCCA AACAAAAATA GGATGTCTGG TTCTGTGCAA AACTAACTCA

        121                                                         180
Ugd10   TCCAGCTATA TAATTCCTTG TCCAACTACA CTGAATTACA TATTTCTTTT GTCCACTTCA
BC-1    TCCAGCTATA TAATTCCTTG TCCAACTACA CTTAATTACA TATTTCTTTT GTCCACTTCA
Cons    TCCAGCTATA TAATTCCTTG TCCAACTACA CT-AATTACA TATTTCTTTT GTCCACTTCA

        181                          start                          240
Ugd10   TTTTTGGGCC TTGGGCTTTT TGTCTACAAT GAATTACAAA AAATACCTGT GGGGTACTTG
BC-1    TTTTTGGGCC TTGGGCTTTT TGTCTACAAT GAATTACAAA AAATACCTGT GGGGTACTTG
Cons    TTTTTGGGCC TTGGGCTTTT TGTCTACAAT GAATTACAAA AAATACCTGT GGGGTACTTG
                                 <-----------------------------------

        241                                                         300
Ugd10   GTTTGCAGCA CTTATAACCT GTTGTGGGTG TTTGTCTATT ATGTTTTGTC TTTTAACTAT
BC-1    GTTTGCAGCA CTTATAACAT GTTGTGGGTG TTTGTCTATT ATGTTTTGTC TTTTAACTAT
Cons    GTTTGCAGCA CTTATAAC-T GTTGTGGGTG TTTGTCTATT ATGTTTTGTC TTTTAACTAT
        ------------------------------------------------------------

        301                                                         360
Ugd10   AAATTTAGAA AACACCATAT TTTTGCTTAG TAACATAAGT GTTTACTACC AACTTTTTTG
BC-1    AAATTTACAA AACACCATAT TTTTGCTCAG TAACATAAGT GTTTACTACC AACTTTTTTG
Cons    AAATTTA-AA AACACCATAT TTTTGCT-AG TAACATAAGT GTTTACTACC AACTTTTTTG
        ------------------------------------------------------------

        361                                                         420
Ugd10   TACAATAACA AACATATATG TGCAGTCAAA AAAACAACGT TTTCAAGCTT CCCCTCCACT
BC-1    TACAATAACA AACATATATG TGCAGTCAAA AAAACAACGT TTTCAAGCTT CCCCTCCCAT
Cons    TACAATAACA AACATATATG TGCAGTCAAA AAAACAACGT TTTCAAGCTT CCCCTCC--T
        ---------*--------------------------------------------------

        421                                                         480
Ugd10   TGGACCATCT ATAATTGTTG GTAAGTTACA TTTTTTAAAT GTTTAACTTA TTTCTTTCTA
BC-1    TGGACCATCT ATAATTGTTG GTAAGTTACA TTTTTTAAAT GTTTAACTTA TTTCTTTCTA
Cons    TGGACCATCT ATAATTGTTG GTAAGTTACA TTTTTTAAAT GTTTAACTTA TTTCTTTCTA
        ---------------------->

        481                                                         540
Ugd10   AGTATACATT GTTTAAATGT CTCATGTATT TGTCTTTTGT TTTAGGATGT ATTGCCTTTG
BC-1    AGTATACATT GTTTAAATGT CTCGTGTATT TGTCTTTTGT TTTAGGATGT ATTGCCTTTG
Cons    AGTATACATT GTTTAAATGT CTC-TGTATT TGTCTTTTGT TTTAGGATGT ATTGCCTTTG
                                                   <---------------

        541                                                         600
Ugd10   CCAGTTGGAG TTTCTCTACT CAAAGTACTT TGAGTACTGT TTGTGTTTGT ATTATAAGTT
BC-1    CCAGTTGGAG TTTCTCTACT CAAAGTACTT TGAGTACTGT TTGTGTTTGT ATTATAAGTT
Cons    CCAGTTGGAG TTTCTCTACT CAAAGTACTT TGAGTACTGT TTGTGTTTGT ATTATAAGTT
        ------------------------------------------------------------

        601                                                         660
Ugd10   TATTGTCTAT AATTACTGGT AAGTTTTACT AAAGGTTTTT TATTAAATAA ACATTATGTA
BC-1    TATTGTCTAT AATTACTGGT AAGTTTTACT AAAGGTTTTT TATTAAATAA AAATTATGTA
Cons    TATTGTCTAT AATTACTGGT AAGTTTTACT AAAGGTTTTT TATTAAATAA A-ATTATGTA
        ------------------>

        661                                                         720
Ugd10   TCTAATATTT CCATTTTTTA AAAAAAACGT AATTCCCCTG TATGTGTTTT TCAGCCATTC
BC-1    TCTAATATTT CCAATTTTTT AAAAAAACGT AATTCCCCTG TATGTGTTTT TCAGCCATTC
Cons    TCTAATATTT CCA-TTTTT- AAAAAAACGT AATTCCCCTG TATGTGTTTT TCAGCCATTC
                                                           <-----

        721                                                         780
Ugd10   TATCACTTGG TGGAACATTA AGAGTCGTTA AATGCACTAT TGACAGCGGA TTACTGTGCA
BC-1    TATCACTTGG TGGAACATTA AGAGTCGTTA AATGCACTAT TGATAGCGGA TTACTGTGCA
Cons    TATCACTTGG TGGAACATTA AGAGTCGTTA AATGCACTAT TGA-AGCGGA TTACTGTGCA
        ------------------------------------------------------------
```

```
        781                                                         840
Ugd10   TTGCTATGGT GTTGGTACTT ATATTCTCAA TGGGATTGCA GATTTACAAC AACTGGACAC
BC-1    TTGTTATGGT GTTGGTACTT ATATTCTCAA TGGGATTGCA GATTTACAAC AACTGGACAC
Cons    TTG-TATGGT GTTGGTACTT ATATTCTCAA TGGGATTGCA GATTTACAAC AACTGGACAC
        ------------------------------------------------------------

        841                                                         900
Ugd10   ATTGCCAGTT TTTTTTGCCA TTATGGACAT TGTTGCTTGT TTTTTTTATA CATATCTTTG
BC-1    ATTGCCAGTT TTTTTTGCCA TTATGGACAT TGTTGCTTGT TTTTTTTATA CATATCTTTG
Cons    ATTGCCAGTT TTTTTTGCCA TTATGGACAT TGTTGCTTGT TTTTTTTATA CATATCTTTG
        ------------------------------------------------------------

        901                                                         960
Ugd10   CAACAAATAA TGGGCCTTGT CTCAAGCTCG CTGCATGTGT GTTTGCAATA TGTGGTGGTA
BC-1    CAACAGATAA TGGGCCTTGT CTCAAGCTCG CTGCATGTGT GTTTGCAATA TGTGGTGGTA
Cons    CAACA-ATAA TGGGCCTTGT CTCAAGCTCG CTGCATGTGT GTTTGCAATA TGTGGTGGTA
        ------------------------------------------------------------>

        961                                                        1020
Ugd10   AGTCTGGTTT TTACTGCTAT GCAATTATGT CTGCTTGCAC CTAAGCATTG CAACCAACAA
BC-1    AGTCTGGTTT TTACTGCTAT GCAATTATGT CTGCTTGCAC CTAAGCATTG CAACCAACAA
Cons    AGTCTGGTTT TTACTGCTAT GCAATTATGT CTGCTTGCAC CTAAGCATTG CAACCAACAA

        1021                                                       1080
Ugd10   TATTATTTAT TTTTTACTAC AGGTATACTC AAGGCAACAC CGGCCTTTTT TTGCGTTTCC
BC-1    TATTATTTAT TTTTTACTAC AGGTATACTC AAGGCAACAC CGGCCTTTTT TTGCGTTTCC
Cons    TATTATTTAT TTTTTACTAC AGGTATACTC AAGGCAACAC CGGCCTTTTT TTGCGTTTCC
                             <----------------------------------------

        1081                                                       1140
Ugd10   CATTCATGCC TTTCTGTAAT CACCGCAGGA TGCATAAGCT GCATACATAT TGGTATATAA
BC-1    CATTCATGCC TTTCTGTAAT CATCGCAGGA TGCATAAGCT GCATACATAT TGGTATATGA
Cons    CATTCATGCC TTTCTGTAAT CA-CGCAGGA TGCATAAGCT GCATACATAT TGGTATAT-A
        ------------------------------------------------------------>

        1141                                                       1200
Ugd10   CTACTGTGAA GTTTACATTA TCTTAACTAA ACATTACCAA TACTATTATA ACATTTTGTA
BC-1    CTACTGTGAA GTTTACATTA TCTTAACTAA ACATTACCAA TACTATTATA ACATTTTGTA
Cons    CTACTGTGAA GTTTACATTA TCTTAACTAA ACATTACCAA TACTATTATA ACATTTTGTA

        1201                                                       1260
Ugd10   ATGAATAACT TTATTACAGG TATGACCGGC TTGTTTATAA CAATGAAACG ACATTGGATT
BC-1    ATGAATAACT TTATTACAGG TATGACCGGC TTGTTTATAA CAATGAAACG ACATTGGATT
Cons    ATGAATAACT TTATTACAGG TATGACCGGC TTGTTTATAA CAATGAAACG ACATTGGATT
                          <---------------------------------------------

        1261                                                       1320
Ugd10   GGATCAACTA AGGGGCTTAT GTCATTTTTA CTTTTACAAG GAGGAGTGTT GGTTACACTA
BC-1    GGATCAACTA AGGGGCTTAT GTCATTTTTA CTTTTACAAG GAGGAGTGTT GGTTACACTA
Cons    GGATCAACTA AGGGGCTTAT GTCATTTTTA CTTTTACAAG GAGGAGTGTT GGTTACACTA
        ------------------------------------------------------------

        1321                                                       1380
Ugd10   ACCACAACAA TAGGGATACT GTTTATAAAG CGCGAACAAG ACACCAATAA CGAAGGTATG
BC-1    ACCACAACAA TAGGGATACT GTTTATAAAG CGCGAACAAG ACACCAATAA CGAAGGTATG
Cons    ACCACAACAA TAGGGATACT GTTTATAAAG CGCGAACAAG ACACCAATAA CGAAGGTATG
        ------------------------------------------------------------>

        1381                                                       1440
Ugd10   TTTTTAGCAG TATAGTGGTT TACATTACAT TTTTT.GACA TACTTATTGG GTGTCATAAA
BC-1    TTTTTACCAG TATAGTGGTT TACATTACAT TTTTTCCACA TACTTATTGG GTGTCATAAA
Cons    TTTTTA-CAG TATAGTGGTT TACATTACAT TTTTT--ACA TACTTATTGG GTGTCATAAA

        1441                                                       1500
Ugd10   TTATTACTGT CCATGTTTTT TATTTACAGG GAGTATAACA TTACTAGCAG GCTGTGGCTT
BC-1    TTATTACTGT CCATATTTTT TATTTACAGG GAGTATAACA TTACTAGCAG GCTGTGGCTT
Cons    TTATTACTGT CCAT-TTTTT TATTTACAGG GAGTATAACA TTACTAGCAG GCTGTGGCTT
                                  <-----------------------------------

        1501                                                       1560
Ugd10   TTTATTATAT TGTTTCTTCT GCTGGCAAAG CTTTCACAAA GCTTCACTCT CTGGTGGCTT
BC-1    TTTATTATAT TGTTTCTTCT GCTGGCAAAG CTTTCACAAA GCTTCACTCT CTGGTGGCTT
Cons    TTTATTATAT TGTTTCTTCT GCTGGCAAAG CTTTCACAAA GCTTCACTCT CTGGTGGCTT
        ------------------------------------------------------------
```

```
        1561                                                        1620
Ugd10   CCTGTTTCTA TTTTTGGGTA AGGATTTTTT GTACCTGGGT AGCAGCTCCC TGGTCATAAA
BC-1    CCTGTTTCTA TTTTTGGGTA AGGATTTTTT GTACCTGGGT AGCAGCTCCC TGGTCATAAA
Cons    CCTGTTTCTA TTTTTGGGTA AGGATTTTTT GTACCTGGGT AGCAGCTCCC TGGTCATAAA
        ----------------->
        1621                                                        1680
Ugd10   GCAAACCGAA ACATAACTTT GGTTTATTTC CTTAGCATGG ACATGTGCTG GATGCTGTGT
BC-1    GCAAACCGAA ACATAACTTT GGTTTATTTC CTTAGCATGG ACATGTGCTG GATGCTGTGT
Cons    GCAAACCGAA ACATAACTTT GGTTTATTTC CTTAGCATGG ACATGTGCTG GATGCTGTGT
                                                   <--------------------------
        1681                                                        1740
Ugd10   TAAGTTAGTC CTGCTCTACA CTGACGGTTG GACTACAGGT GTCACTTCAG GACTGATTTG
BC-1    TAAGTTAGTC CTGCTCTACA CTGACGGTTG GACTACAGGT GTTACTTCAG GACTGATTTG
Cons    TAAGTTAGTC CTGCTCTACA CTGACGGTTG GACTACAGGT GT-ACTTCAG GACTGATTTG
        ------------------------------------------------------------------
        1741                                                        1800
Ugd10   TGTAATTGTG ATATTAAGTG AGTTGTGTTT TTATTTATGA TTGAAACACC AGCGCCACAT
BC-1    TGTAATTGTG ATATTAAGTG AGTTGTGTTT TTATTTATAA TTGGAACACC AGCGCCACAT
Cons    TGTAATTGTG ATATTAAGTG AGTTGTGTTT TTATTTAT-A TTG-AACACC AGCGCCACAT
        ----------------->
        1801                                                        1860
Ugd10   AAAAACAGAG ATTAACGATT TTATTTTTTC TTATTAGGTA CTGGCCAAGC CGTACTGGTG
BC-1    AAAAACAGAG ATTAACGATT TTATTTTTTC TTATTAGGTA CTGGCCAAGC CGTACTGGTG
Cons    AAAAACAGAG ATTAACGATT TTATTTTTTC TTATTAGGTA CTGGCCAAGC CGTACTGGTG
                                                   <-------------------------
        1861                                                        1920
Ugd10   GGTTATCTCT ACCGAGAGAG CAGACTTGTG TCGTTCAACA ATGTAACAAC AAGATTACCA
BC-1    GGTTATCTCT ACCGAGAGAG CAGACTTGTG TCGTTCAACA ATGTAACAAC AAGATTACCA
Cons    GGTTATCTCT ACCGAGAGAG CAGACTTGTG TCGTTCAACA ATGTAACAAC AAGATTACCA
        ------------------------------------------------------------------
        1921                                                        1980
Ugd10   ATATATACAC CCCACGACAC ACCACATGCT CATGCAGGCA GGATATGTCC CGATGTTAAT
BC-1    ATATATACAC CACACGACAC ACCACATGCT CATGCAGGCA GGATATGTCC CGATGTTAAT
Cons    ATATATACAC C-CACGACAC ACCACATGCT CATGCAGGCA GGATATGTCC CGATGTTAAT
        ------------------------------------------------------------------
        1981                                                        2040
Ugd10   CATTTAGCTC GCCGTTTACC ACCTCTACCA TCTAGAAATC TTATACACTC GCGTATTTTA
BC-1    CATTTAGCTC GCCGTTTACC ACCTCTACCA TCTAGAAATG TTATACACTC GCGTATTTTA
Cons    CATTTAGCTC GCCGTTTACC ACCTCTACCA TCTAGAAAT- TTATACACTC GCGTATTTTA
        ------------------------------------------------------------------
        2041                                                        2100
Ugd10   AGTTCCACAA CAGACATGGC ATTATCTCCA ATAAGGGTTT GCAACACAGA AGTAACGACC
BC-1    AGTTCCACAA CAGACATGGC ATTATCTCCA GTAAGGGTTT GCAACACAGA AGTAACGACC
Cons    AGTTCCACAA CAGACATGGC ATTATCTCCA -TAAGGGTTT GCAACACAGA AGTAACGACC
        ------------------------------------------------------------------
        2101                                                        2160
Ugd10   CAACTTGAAA TGCAGCAACT ACATAACGAA CACACAGTCA CCTATGCTAG ATTCTTGGGC
BC-1    CAACTTGAAA TGCAGCAACT ACATAGCGAA CGCACAGTCA CCTATGCTAG ATTCTTGGGC
Cons    CAACTTGAAA TGCAGCAACT ACATA-CGAA C-CACAGTCA CCTATGCTAG ATTCTTGGGC
        ------------------------------------------------------------------
        2161                                                        2220
Ugd10   GACAACACGC CTCCACCAAC GCGTGCGTCC GCTTGTATTA ACCAATCAGG TATCTCTAAT
BC-1    GACAACACGC CTCCACCAAC GCGTGCGTCC GCTTGTATTA ACCAATCAGG TATCTCTAAT
Cons    GACAACACGC CTCCACCAAC GCGTGCGTCC GCTTGTATTA ACCAATCAGG TATCTCTAAT
        ------------------------------------------------------------------
        2221                                                        2280
Ugd10   GTGAGCAACT GTGGTGTAAG GAGTCTGGAT CCGCCACCAT TTCAGCCTGC AGATGAAGTG
BC-1    GTGAGCAACT GTGGTGTAAG GAGTCTGGAT CCGCCACCAT TTCAGCCTGC AGATGAAGTG
Cons    GTGAGCAACT GTGGTGTAAG GAGTCTGGAT CCGCCACCAT TTCAGCCTGC AGATGAAGTG
        ------------------------------------------------------------------
        2281                          stop                          2340
Ugd10   TATGAGGAAG TTTTGTTTCC CACGGACTAA CCTAGGACCA CAGCATTTTT CCATCGATAG
BC-1    TATGAGGAAG TTTTGTTTCC CACGGACTAA CCTAGGACCA CAGCATTTTT CCATCGATAG
Cons    TATGAGGAAG TTTTGTTTCC CACGGACTAA CCTAGGACCA CAGCATTTTT CCATCGATAG
        ---------------------------------> <        21 bp        >
```

```
       2341                                                        2400
Ugd10  CTTTTTTTGA GTATTTGAGG TTAGTGACAT GGCTACAAGT AACTGTGGAT CCCATGAAAG
BC-1   CTTTTTTTGA GTATTTGAGG TTAGTGACAT GGCTACAAGT AACTGTGGAT CCCATGAAAG
Cons   CTTTTTTTGA GTATTTGAGG TTAGTGACAT GGCTACAAGT AACTGTGGAT CCCATGAAAG

       2401                                                        2460
Ugd10  CGGAAACAGA CTCTGCCCGA CAACCTGTGA CTGCGACATA TTTGGGGATT CTAAGTTTCT
BC-1   CGGAAACAGA CTCTGCCCGA CAACCTGTGA CTGCGACATA TTTGGGGATT CTAAGTTTCT
Cons   CGGAAACAGA CTCTGCCCGA CAACCTGTGA CTGCGACATA TTTGGGGATT CTAAGTTTCT

       2461                                                        2520
Ugd10  TCCTACTCCC CAGAGATCTC GCGGGTTTCT CGGCAGCCTG ACTACAGAGG GTGTCCCCGG
BC-1   TCCTACTCCC CAGAGATCTC GCGGGTTTCT CGGCAGCCTG ACTACAGAGG GTGTCCCCGG
Cons   TCCTACTCCC CAGAGATCTC GCGGGTTTCT CGGCAGCCTG ACTACAGAGG GTGTCCCCGG

       2521             2540
Ugd10  GGGCGGTGCG CCCTCTAGGC
BC-1   GGGCGGTGCG CCCTCTAGGC
Cons   GGGCGGTGCG CCCTCTAGGC
```

Fig. 5.4. Nucleotide changes within (A) HHV-8 K15 M and (B) K15 P alleles. Sequences of Ugandan strains (bold) were analysed together with those of BC-1, BCBL-R and GK18 (Table 2). K1 subtypes are indicated. The results are presented in the context of variation from the BC-1 (A) or BCBL-R (B) sequences, hyphens denoting identity. The alignments in each case consist of the entire K15 gene oriented from left to right (opposite to the genomic orientation; nucleotides 1 and 2102 in K15 M are equivalent to 136772 and 134672, respectively, in the BC-1 genome). The BC-1 sequence was modified before alignment by replacing the missing T residue at 136610 (see text). Nucleotides 1 and 2086 in K15 P are equivalent to 3652 and 1578, respectively, in the BCBL-R sequence and to 21490 and 19415, respectively, in the GK18 sequence. The position in the final alignment is indicated in the 'nucleotide' row. The 'exon/intron' row shows the number of the exon (E1-8) or intron (I1-7) in which each substitution or deletion is located. Amino acid differences are depicted in the 'residue' row, asterisks and exclamation marks indicating synonymous substitutions in exons and non-coding substitutions in introns, respectively. Deletions (Del) are represented by dots.

A. K15 M (2102 bp)

nucleotide		51	100	120	210	211	296	444	466	472	556	576	698
exon/intron		E1	E1	E1	E1	E1	I1	I2	I2	I2	E3	E3	E3
residue	K1 subtype	*	Q/E	*	*	I/L	!	!	!	!	*	V/A	D/N
BC-1	A	ACA	CAA	CTC	CCC	ATT	G	A	A	T	GAT	GTT	GAT
Ugd10	**B**	--C	G--	--T	--A	C--	A	C	T	A	--C	-C-	A--

nucleotide		895	931	1179	1208	1209	1247	1515	1571	1576	1724	1812	1863	1918	1924
exon/intron		E4	I4	I5	I5	I5	I5	E7	I7	I7	E8	E8	E8	E8	E8
residue	K1 subtype	I/T	!	!	Del	!	!	*	!	!	*	V/L	V/I	S/N	R/H
BC-1	A	ATC	G	C	C	C	A	GTT	A	G	CCA	GTT	GTA	AGC	CGC
Ugd10	**B**	-C-	A	G	.	G	G	--C	G	A	--C	C--	A--	-A-	-A-

B. K15 P (2086 bp)

nucleotide		6	8	11	14	16-18	45	223-232	303	341	366	437	454	521	576	589	601	607	655
exon/intron		E1	E1	E1	E1	E1	E1	I1	I1	E2	E2	I2	I2	E3	E3	E3	E3	E3	E3
residue	K1 subtype	K/N	T/I	L/H	I/R	Del	*	Del	!	*	S/T	!	!	T/A	F/C	*	M/I	*	*
BCBL-R	A	AAG	ACA	CTC	ATA	TTC	CTA	....	A	TGC	TCA	G	C	ACG	TTC	TTT	ATG	AAC	TTT
GK18	C	---	---	---	---	---	---	....	-	---	---	-	G	---	---	---	---	---	---
Ugd4	**A5**	---	-T-	---	-G-	---	---	....	-	--T	---	-	-	---	---	---	---	---	---
Ugd16	**A5**	---	-T-	---	-G-	---	---	....	-	--T	---	-	-	---	---	---	---	---	---
Ugd15	**B**	---	---	---	---	---	---	....	-	--T	---	-	-	---	---	--C	---	---	--C
Ugd23	**C**	--C	---	---	---	---	---	....	-	--T	---	-	-	---	---	--C	---	---	--C
Ugd29	**B**	---	---	---	---	...	---	....	-	--T	---	-	-	---	---	---	---	---	---
Ugd2	**B**	---	---	-A-	---	---	--G	TAGACTTTCA	C	--T	A--	A	-	G--	-G-	---	--A	--T	---
Ugd12	**A5**	---	---	-A-	---	---	--G	TAGACTTTCA	C	--T	A--	A	-	G--	-G-	---	--A	--T	---
Ugd19	**B**	---	---	-A-	---	---	--G	TAGACTTTCA	C	--T	A--	A	-	G--	-G-	--	--A	--T	---

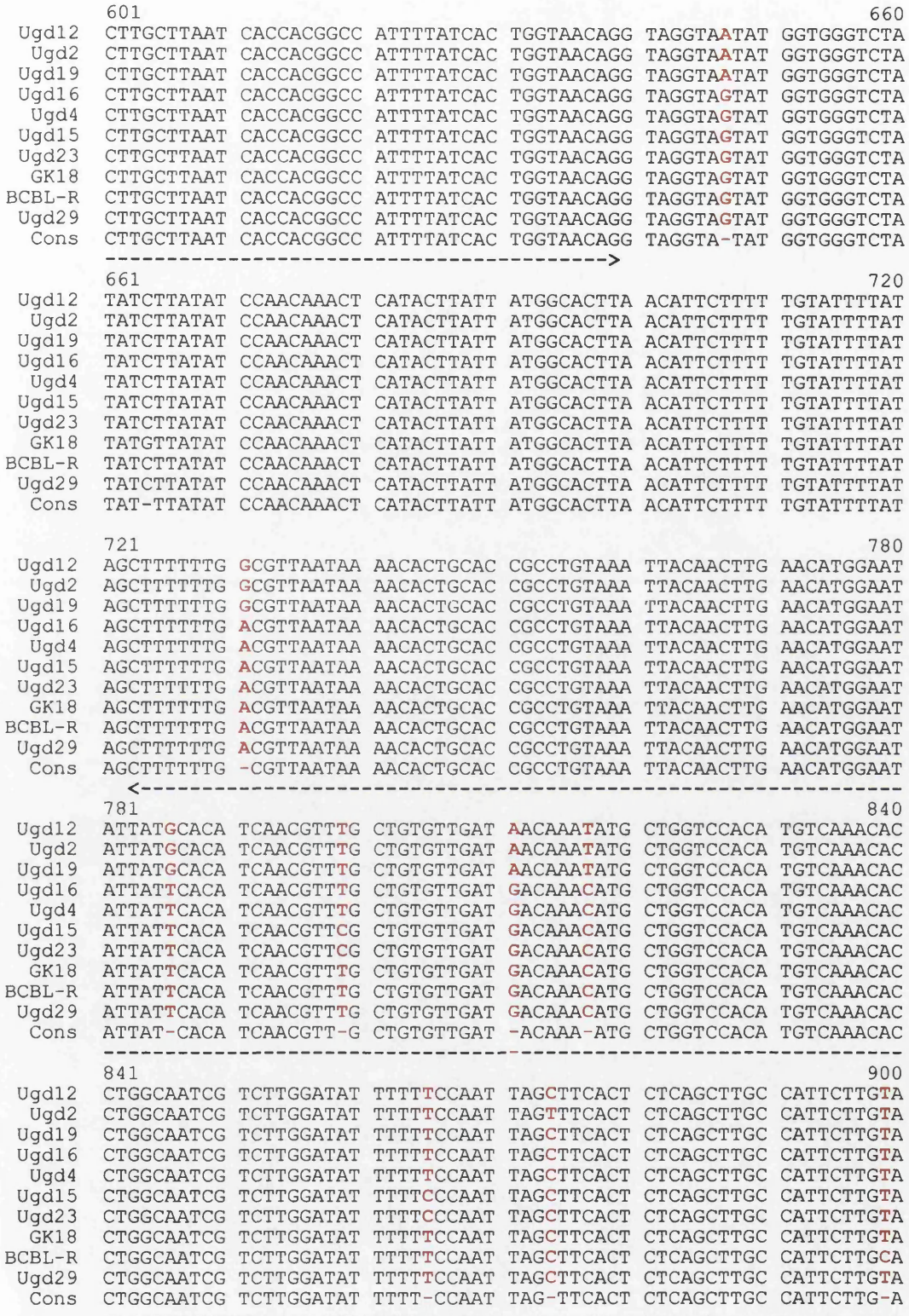
nucleotide		664	689	791	833	895	962	974	1029	1135	1211	1317	1337	1403	1599	1627	1668	1673	1851	2023
exon/intron		E3	E3	I3	I3	E4	I4	I4	E5	E5	I5	E6	E6	I6	I7	E8	E8	E8	E8	E8
residue	K1 subtype	*	H/Y	!	Del	A/V	!	!	S/T	*	!	A/T	M/I	!	!	*	K/T	V/L	P/L	*
BCBL-R	A	AGC	CAT	A	.	GCC	G	G	AGC	GCA	G	GCG	ATG	G	G	ACT	AAA	GTG	CCA	ACC
GK18	C	---	T--	-	T	---	-	-	---	---	T	---	---	-	-	--G	---	---	-T-	---
Ugd4	**A5**	---	T--	-	T	---	-	-	---	---	-	---	---	-	-	--G	---	---	---	---
Ugd16	**A5**	---	T--	-	T	---	-	-	---	---	-	---	---	-	-	--G	-C-	---	---	--T
Ugd15	**B**	---	T--	-	T	---	-	-	---	---	-	---	---	-	-	--G	---	---	---	---
Ugd23	**C**	---	T--	-	T	---	-	-	---	---	-	---	---	-	-	--G	---	---	---	---
Ugd29	**B**	---	T--	-	T	---	-	-	---	--C	-	---	---	-	-	--G	---	T--	---	---
Ugd2	**B**	--T	T--	G	T	-T-	-	A	---	---	-	---	--A	T	-	--G	---	---	---	---
Ugd12	**A5**	---	T--	G	T	---	A	A	---	---	-	A--	--A	T	-	--G	---	---	---	---
Ugd19	**B**	---	T--	G	T	---	-	A	-C-	---	-	---	--A	T	C	--G	---	---	---	---

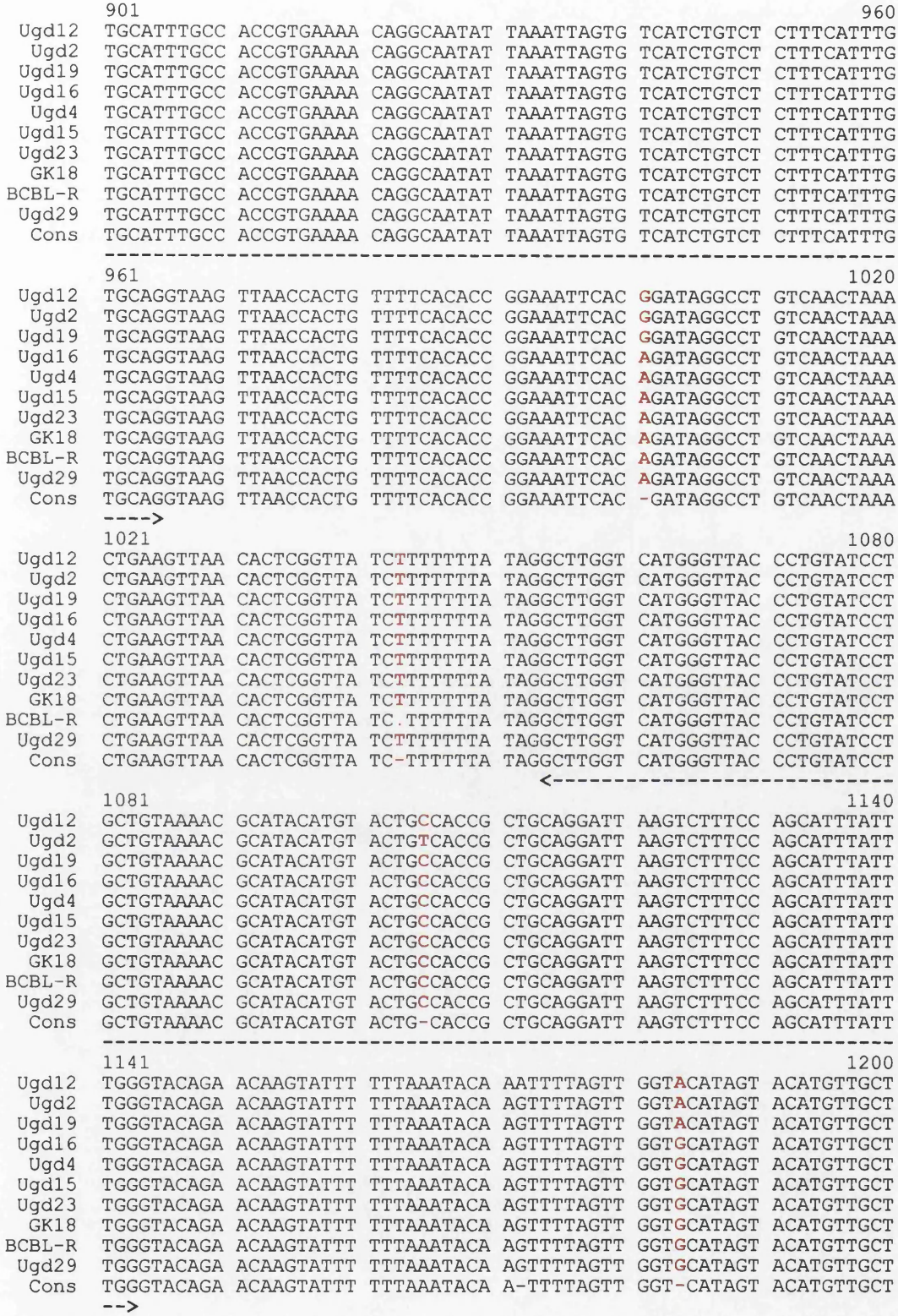
Fig. 5.5. DNA alignment of K15 P allele sequences.

The alignment includes the entire K15 gene, and sequences (approx. 200 bp) flanking both sides, of strains with the P allele. Sequences of Ugandan strains were analysed together with those of BCBL-R and GK18. The alignment is oriented from left to right, i.e. opposite to the genomic orientation. Nucleotides 1 and 2505 are equivalent to 3862 and 1369, respectively, in the BCBL-R sequence and to 21700 and 19161, respectively, in the GK18 sequence. The start and stop codons are highlighted in bold-blue. Sequences in exons are underlined with a dashed line enclosed within '<' and '>' signs. Positions of nucleotide changes are shown in bold-red, with deletions represented by dots. The position (nucleotide 2321) where the M and P alleles become distinct is highlighted in bold-pink.

```
          1                                                          60
Ugd12     ACAATTTACG AGCCTTGTAT CCGGAATATT TATGAGCCTT GTGTCGGGAA TACTTAGGAA
Ugd2      ACAATTTACG AGCCTTGTAT CCGGAATATT TATGAGCCTT GTGTCGGGAA TACTTAGGAA
Ugd19     ACAATTTACG AGCCTTGTAT CCGGAATATT TATGAGCCTT GTGTCGGGAA TACTTAGGAA
Ugd16     ACAATTTACG AGCCTTGTAT CCGGAATATT TATGAGCCTT GTGTCGGGAA TACTTAGGAA
Ugd4      ACAATTTACG AGCCTTGTAT CCGGAATATT TATGAGCCTT GTGTCGGGAA TACTTAGGAA
Ugd15     ACAATTTACG AGCCTTGTAT CCGGAATATT TATGAGCCTT GTGTCGGGAA TACTTAGGAA
Ugd23     ACAATTTACG AGCCTTGTAT CCGGAATATT TATGAGCCTT GTGTCGGGAA TACTTAGGAA
GK18      ACAATTTACG AGCCTTGTAT CCGGAATATT TATGAGCCTT GTGTCGGGAA TACTTAGGAA
BCBL-R    ACAATTTACG AGCCTTGTAT CCGGAATATT TATGAGCCTT GTGTCGGGAA TACTTAGGAA
Ugd29     ACAATTTACG AGCCTTGTAT CCGGAATATT TATGAGCCTT GTGTCGGGAA TACTTAGGAA
Cons      ACAATTTACG AGCCTTGTAT CCGGAATATT TATGAGCCTT GTGTCGGGAA TACTTAGGAA

          61                                                        120
Ugd12     ATCAGCAAAA TTCCATGTGG CTTCCTAAAA CAATTGTACA CAAAACCTTA AATGCCCTTT
Ugd2      ATCAGCAAAA TTCCATGTGG CTTCCTAAAA CAATTGTACA CAAAACCTTA AATGCCCTTT
Ugd19     ATCAGCAAAA TTCCATGTGG CTTCCTAAAA CAATTGTACA CAAAACCTTA AATGCCCTTT
Ugd16     ATCAGCAAAA TTCCATGTGG CTTCCTAAAA CAATTGTACA CAAAACCTTA AATGCCCTTT
Ugd4      ATCAGCAAAA TTCCATGTGG CTTCCTAAAA CAATTGTACA CAAAACCTTA AATGCCCTTT
Ugd15     ATCAGCAAAA TTCCATGTGG CTTCCTAAAA CAATTGTACA CAAAACCTTA AATGCCCTTT
Ugd23     ATCAGCAAAA TTCCATGTGG CTTCCTAAAA CAATTGTACA CAAAACCTTA AATGCCCTTT
GK18      ATCAGCAAAA TTCCATGTGG CTTCCTAAAA CAATTGTACA CAAAACCTTA AATGCCCTTT
BCBL-R    ATCAGCAAAA TTCCATGTGG CTTCCTAAAA CAATTGTACA CAAAACCTTA AATGCCCTTT
Ugd29     ATCAGCAAAA TTCCATGTGG CTTCCTAAAA CAATTGTACA CAAAACCTTA AATGCCCTTT
Cons      ATCAGCAAAA TTCCATGTGG CTTCCTAAAA CAATTGTACA CAAAACCTTA AATGCCCTTT

          121                                                       180
Ugd12     CTAGTATCAC GTGAATCCTA AAACGTATTT AAAAATCACA AACCCCATTT ACAACAACTC
Ugd2      CTAGTATCAC GTGAATCCTA AAACGTATTT AAAAATCACA AACCCCATTT ACAACAACTC
Ugd19     CTAGTATCAC GTGAATCCTA AAACGTATTT AAAAATCACA AACCCCATTT ACAACAACTC
Ugd16     CTAGTATCAC GTGAATCCTA AAACGTATTT AAAAATCACA AACCCCATTT ACAACAACTC
Ugd4      CTAGTATCAC GTGAATCCTA AAACGTATTT AAAAATCACA AACCCCATTT ACAACAACTC
Ugd15     CTAGTATCAC GTGAATCCTA AAACGTATTT AAAAATCACA AACCCCATTT ACAACAACTC
Ugd23     CTAGTATCAC GTGAATCCTA AAACGTATTT AAAAATCACA AACCCCATTT ACAACAACTC
GK18      CTAGTATCAC GTGAATCCTA AAACGTATTT AAAAATCACA AACCCCATTT ACAACAACTC
BCBL-R    CTAGTATCAC GTGAATCCTA AAACGTATTT AAAAATCACA AACCCCATTT ACAACAACTC
Ugd29     CTAGTATCAC GTGAATCCTA AAACGTATTT AAAAATCACA AACCCCATTT ACAACAACTC
Cons      CTAGTATCAC GTGAATCCTA AAACGTATTT AAAAATCACA AACCCCATTT ACAACAACTC

          181                          start                        240
Ugd12     TATTGTAAGC CCTGTGGATA CCTAGTCAAA ATGAAGACAC ACATATTCTT CTGGAATTTA
Ugd2      TATTGTAAGC CCTGTGGATA CCTAGTCAAA ATGAAGACAC ACATATTCTT CTGGAATTTA
Ugd19     TATTGTAAGC CCTGTGGATA CCTAGTCAAA ATGAAGACAC ACATATTCTT CTGGAATTTA
Ugd16     TATTGTAAGC CCTGTGGATA CCTAGTCAAA ATGAAGATAC TCAGATTCTT CTGGAATTTA
Ugd4      TATTGTAAGC CCTGTGGATA CCTAGTCAAA ATGAAGATAC TCAGATTCTT CTGGAATTTA
Ugd15     TATTGTAAGC CCTGTGGATA CCTAGTCAAA ATGAAGACAC TCATATTCTT CTGGAATTTA
Ugd23     TATTGTAAGC CCTGTGGATA CCTAGTCAAA ATGAACACAC TCATATTCTT CTGGAATTTA
GK18      TATTGTAAGC CCTGTGGATA CCTAGTCAAA ATGAAGACAC TCATATTCTT CTGGAATTTA
BCBL-R    TATTGTAAGC CCTGTGGATA CCTAGTCAAA ATGAAGACAC TCATATTCTT CTGGAATTTA
Ugd29     TATTGTAAGC CCTGTGGATA CCTAGTCAAA ATGAAGACAC TCATA...TT CTGGAATTTA
Cons      TATTGTAAGC CCTGTGGATA CCTAGTCAAA ATGAA-A-AC -CA-A---TT CTGGAATTTA
                                           <------------------------------

          241                                                       300
Ugd12     TGGCTTTGGG CCCTGCTGGT ATGTTTTTGG TGTATCACTC TTGTCTGTGT AACTACCAAC
Ugd2      TGGCTTTGGG CCCTGCTGGT ATGTTTTTGG TGTATCACTC TTGTCTGTGT AACTACCAAC
Ugd19     TGGCTTTGGG CCCTGCTGGT ATGTTTTTGG TGTATCACTC TTGTCTGTGT AACTACCAAC
Ugd16     TGGCTTTGGG CCCTACTGGT ATGTTTTTGG TGTATCACTC TTGTCTGTGT AACTACCAAC
Ugd4      TGGCTTTGGG CCCTACTGGT ATGTTTTTGG TGTATCACTC TTGTCTGTGT AACTACCAAC
Ugd15     TGGCTTTGGG CCCTACTGGT ATGTTTTTGG TGTATCACTC TTGTCTGTGT AACTACCAAC
Ugd23     TGGCTTTGGG CCCTACTGGT ATGTTTTTGG TGTATCACTC TTGTCTGTGT AACTACCAAC
GK18      TGGCTTTGGG CCCTACTGGT ATGTTTTTGG TGTATCACTC TTGTCTGTGT AACTACCAAC
BCBL-R    TGGCTTTGGG CCCTACTGGT ATGTTTTTGG TGTATCACTC TTGTCTGTGT AACTACCAAC
Ugd29     TGGCTTTGGG CCCTACTGGT ATGTTTTTGG TGTATCACTC TTGTCTGTGT AACTACCAAC
Cons      TGGCTTTGGG CCCT-CTGGT ATGTTTTTGG TGTATCACTC TTGTCTGTGT AACTACCAAC
          ------------------------------------------------------------
```

```
        301                                                         360
Ugd12   TCAATTGATA CAATGGCTTC TTTGCTTGTT ATGTGCATTT TGTTTGTCAG TGCTATTAAT
 Ugd2   TCAATTGATA CAATGGCTTC TTTGCTTGTT ATGTGCATTT TGTTTGTCAG TGCTATTAAT
Ugd19   TCAATTGATA CAATGGCTTC TTTGCTTGTT ATGTGCATTT TGTTTGTCAG TGCTATTAAT
Ugd16   TCAATTGATA CAATGGCTTC TTTGCTTGTT ATGTGCATTT TGTTTGTCAG TGCTATTAAT
 Ugd4   TCAATTGATA CAATGGCTTC TTTGCTTGTT ATGTGCATTT TGTTTGTCAG TGCTATTAAT
Ugd15   TCAATTGATA CAATGGCTTC TTTGCTTGTT ATGTGCATTT TGTTTGTCAG TGCTATTAAT
Ugd23   TCAATTGATA CAATGGCTTC TTTGCTTGTT ATGTGCATTT TGTTTGTCAG TGCTATTAAT
 GK18   TCAATTGATA CAATGGCTTC TTTGCTTGTT ATGTGCATTT TGTTTGTCAG TGCTATTAAT
BCBL-R  TCAATTGATA CAATGGCTTC TTTGCTTGTT ATGTGCATTT TGTTTGTCAG TGCTATTAAT
Ugd29   TCAATTGATA CAATGGCTTC TTTGCTTGTT ATGTGCATTT TGTTTGTCAG TGCTATTAAT
 Cons   TCAATTGATA CAATGGCTTC TTTGCTTGTT ATGTGCATTT TGTTTGTCAG TGCTATTAAT
        ------------------------------------------------------------

        361                                                         420
Ugd12   AAGTATACAC AGGCAATCTC TAGCAACAAT CCTAAATGGC CTTCATCCTG GCACCTAGGA
 Ugd2   AAGTATACAC AGGCAATCTC TAGCAACAAT CCTAAATGGC CTTCATCCTG GCACCTAGGA
Ugd19   AAGTATACAC AGGCAATCTC TAGCAACAAT CCTAAATGGC CTTCATCCTG GCACCTAGGA
Ugd16   AAGTATACAC AGGCAATCTC TAGCAACAAT CCTAAATGGC CTTCATCCTG GCACCTAGGA
 Ugd4   AAGTATACAC AGGCAATCTC TAGCAACAAT CCTAAATGGC CTTCATCCTG GCACCTAGGA
Ugd15   AAGTATACAC AGGCAATCTC TAGCAACAAT CCTAAATGGC CTTCATCCTG GCACCTAGGA
Ugd23   AAGTATACAC AGGCAATCTC TAGCAACAAT CCTAAATGGC CTTCATCCTG GCACCTAGGA
 GK18   AAGTATACAC AGGCAATCTC TAGCAACAAT CCTAAATGGC CTTCATCCTG GCACCTAGGA
BCBL-R  AAGTATACAC AGGCAATCTC TAGCAACAAT CCTAAATGGC CTTCATCCTG GCACCTAGGA
Ugd29   AAGTATACAC AGGCAATCTC TAGCAACAAT CCTAAATGGC CTTCATCCTG GCACCTAGGA
 Cons   AAGTATACAC AGGCAATCTC TAGCAACAAT CCTAAATGGC CTTCATCCTG GCACCTAGGA
        ------------------------------------------------------------

        421                                                         480
Ugd12   ATTATTGGTA AGTAGACTTT CATACTGGCT ACCAGGAATT CCTGGTAAGT AGCTTTTCCT
 Ugd2   ATTATTGGTA AGTAGACTTT CATACTGGCT ACCAGGAATT CCTGGTAAGT AGCTTTTCCT
Ugd19   ATTATTGGTA AGTAGACTTT CATACTGGCT ACCAGGAATT CCTGGTAAGT AGCTTTTCCT
Ugd16   ATTATTGGTA AG........ ..TACTGGCT ACCAGGAATT CCTGGTAAGT AGCTTTTCCT
 Ugd4   ATTATTGGTA AG........ ..TACTGGCT ACCAGGAATT CCTGGTAAGT AGCTTTTCCT
Ugd15   ATTATTGGTA AG........ ..TACTGGCT ACCAGGAATT CCTGGTAAGT AGCTTTTCCT
Ugd23   ATTATTGGTA AG........ ..TACTGGCT ACCAGGAATT CCTGGTAAGT AGCTTTTCCT
 GK18   ATTATTGGTA AG........ ..TACTGGCT ACCAGGAATT CCTGGTAAGT AGCTTTTCCT
BCBL-R  ATTATTGGTA AG........ ..TACTGGCT ACCAGGAATT CCTGGTAAGT AGCTTTTCCT
Ugd29   ATTATTGGTA AG........ ..TACTGGCT ACCAGGAATT CCTGGTAAGT AGCTTTTCCT
 Cons   ATTATTGGTA AG-------- --TACTGGCT ACCAGGAATT CCTGGTAAGT AGCTTTTCCT
        ------>
        481                                                         540
Ugd12   TTTTTTCATT GTAGTATAAA CATTGCAAGC TGCCACTTGT ATATTCACAT ACATTTGTTT
 Ugd2   TTTTTTCATT GTAGTATAAA CATTGCAAGC TGCCACTTGT ATATTCACAT ACATTTGTTT
Ugd19   TTTTTTCATT GTAGTATAAA CATTGCAAGC TGCCACTTGT ATATTCACAT ACATTTGTTT
Ugd16   TTTTTTCATT GTAGTATAAA CATTGCAAGC TGACACTTGT ATATTCACAT ACATTTGTTT
 Ugd4   TTTTTTCATT GTAGTATAAA CATTGCAAGC TGACACTTGT ATATTCACAT ACATTTGTTT
Ugd15   TTTTTTCATT GTAGTATAAA CATTGCAAGC TGACACTTGT ATATTCACAT ACATTTGTTT
Ugd23   TTTTTTCATT GTAGTATAAA CATTGCAAGC TGACACTTGT ATATTCACAT ACATTTGTTT
 GK18   TTTTTTCATT GTAGTATAAA CATTGCAAGC TGACACTTGT ATATTCACAT ACATTTGTTT
BCBL-R  TTTTTTCATT GTAGTATAAA CATTGCAAGC TGACACTTGT ATATTCACAT ACATTTGTTT
Ugd29   TTTTTTCATT GTAGTATAAA CATTGCAAGC TGACACTTGT ATATTCACAT ACATTTGTTT
 Cons   TTTTTTCATT GTAGTATAAA CATTGCAAGC TG-CACTTGT ATATTCACAT ACATTTGTTT

        541                                                         600
Ugd12   TTATAGCTTG TATTGTCCTT AAACTTTGGA ACTTGACAAC TACCAACTCT GTAACCTATG
 Ugd2   TTATAGCTTG TATTGTCCTT AAACTTTGGA ACTTGACAAC TACCAACTCT GTAACCTATG
Ugd19   TTATAGCTTG TATTGTCCTT AAACTTTGGA ACTTGACAAC TACCAACTCT GTAACCTATG
Ugd16   TTATAGCTTG TATTGTCCTT AAACTTTGGA ACTTGTCAAC TACCAACTCT GTAACCTATG
 Ugd4   TTATAGCTTG TATTGTCCTT AAACTTTGGA ACTTGTCAAC TACCAACTCT GTAACCTATG
Ugd15   TTATAGCTTG TATTGTCCTT AAACTTTGGA ACTTGTCAAC TACCAACTCT GTAACCTATG
Ugd23   TTATAGCTTG TATTGTCCTT AAACTTTGGA ACTTGTCAAC TACCAACTCT GTAACCTATG
 GK18   TTATAGCTTG CATTGTCCTT AAACTTTGGA ACTTGTCAAC TACCAACTCT GTAACCTATG
BCBL-R  TTATAGCTTG CATTGTCCTT AAACTTTGGA ACTTGTCAAC TACCAACTCT GTAACCTATG
Ugd29   TTATAGCTTG TATTGTCCTT AAACTTTGGA ACTTGTCAAC TACCAACTCT GTAACCTATG
 Cons   TTATAGCTTG -ATTGTCCTT AAACTTTGGA ACTTG-CAAC TACCAACTCT GTAACCTATG
           <------------------------------------------------------------
```


601 660
Ugd12 CTTGCTTAAT CACCACGGCC ATTTTATCAC TGGTAACAGG TAGGTAATAT GGTGGGTCTA
Ugd2 CTTGCTTAAT CACCACGGCC ATTTTATCAC TGGTAACAGG TAGGTAATAT GGTGGGTCTA
Ugd19 CTTGCTTAAT CACCACGGCC ATTTTATCAC TGGTAACAGG TAGGTAATAT GGTGGGTCTA
Ugd16 CTTGCTTAAT CACCACGGCC ATTTTATCAC TGGTAACAGG TAGGTAGTAT GGTGGGTCTA
Ugd4 CTTGCTTAAT CACCACGGCC ATTTTATCAC TGGTAACAGG TAGGTAGTAT GGTGGGTCTA
Ugd15 CTTGCTTAAT CACCACGGCC ATTTTATCAC TGGTAACAGG TAGGTAGTAT GGTGGGTCTA
Ugd23 CTTGCTTAAT CACCACGGCC ATTTTATCAC TGGTAACAGG TAGGTAGTAT GGTGGGTCTA
GK18 CTTGCTTAAT CACCACGGCC ATTTTATCAC TGGTAACAGG TAGGTAGTAT GGTGGGTCTA
BCBL-R CTTGCTTAAT CACCACGGCC ATTTTATCAC TGGTAACAGG TAGGTAGTAT GGTGGGTCTA
Ugd29 CTTGCTTAAT CACCACGGCC ATTTTATCAC TGGTAACAGG TAGGTAGTAT GGTGGGTCTA
Cons CTTGCTTAAT CACCACGGCC ATTTTATCAC TGGTAACAGG TAGGTA-TAT GGTGGGTCTA
661 720
Ugd12 TATCTTATAT CCAACAAACT CATACTTATT ATGGCACTTA ACATTCTTTT TGTATTTTAT
Ugd2 TATCTTATAT CCAACAAACT CATACTTATT ATGGCACTTA ACATTCTTTT TGTATTTTAT
Ugd19 TATCTTATAT CCAACAAACT CATACTTATT ATGGCACTTA ACATTCTTTT TGTATTTTAT
Ugd16 TATCTTATAT CCAACAAACT CATACTTATT ATGGCACTTA ACATTCTTTT TGTATTTTAT
Ugd4 TATCTTATAT CCAACAAACT CATACTTATT ATGGCACTTA ACATTCTTTT TGTATTTTAT
Ugd15 TATCTTATAT CCAACAAACT CATACTTATT ATGGCACTTA ACATTCTTTT TGTATTTTAT
Ugd23 TATCTTATAT CCAACAAACT CATACTTATT ATGGCACTTA ACATTCTTTT TGTATTTTAT
GK18 TATGTTATAT CCAACAAACT CATACTTATT ATGGCACTTA ACATTCTTTT TGTATTTTAT
BCBL-R TATCTTATAT CCAACAAACT CATACTTATT ATGGCACTTA ACATTCTTTT TGTATTTTAT
Ugd29 TATCTTATAT CCAACAAACT CATACTTATT ATGGCACTTA ACATTCTTTT TGTATTTTAT
Cons TAT-TTATAT CCAACAAACT CATACTTATT ATGGCACTTA ACATTCTTTT TGTATTTTAT
721 780
Ugd12 AGCTTTTTTG GCGTTAATAA AACACTGCAC CGCCTGTAAA TTACAACTTG AACATGGAAT
Ugd2 AGCTTTTTTG GCGTTAATAA AACACTGCAC CGCCTGTAAA TTACAACTTG AACATGGAAT
Ugd19 AGCTTTTTTG GCGTTAATAA AACACTGCAC CGCCTGTAAA TTACAACTTG AACATGGAAT
Ugd16 AGCTTTTTTG ACGTTAATAA AACACTGCAC CGCCTGTAAA TTACAACTTG AACATGGAAT
Ugd4 AGCTTTTTTG ACGTTAATAA AACACTGCAC CGCCTGTAAA TTACAACTTG AACATGGAAT
Ugd15 AGCTTTTTTG ACGTTAATAA AACACTGCAC CGCCTGTAAA TTACAACTTG AACATGGAAT
Ugd23 AGCTTTTTTG ACGTTAATAA AACACTGCAC CGCCTGTAAA TTACAACTTG AACATGGAAT
GK18 AGCTTTTTTG ACGTTAATAA AACACTGCAC CGCCTGTAAA TTACAACTTG AACATGGAAT
BCBL-R AGCTTTTTTG ACGTTAATAA AACACTGCAC CGCCTGTAAA TTACAACTTG AACATGGAAT
Ugd29 AGCTTTTTTG ACGTTAATAA AACACTGCAC CGCCTGTAAA TTACAACTTG AACATGGAAT
Cons AGCTTTTTTG -CGTTAATAA AACACTGCAC CGCCTGTAAA TTACAACTTG AACATGGAAT
781 840
Ugd12 ATTATGCACA TCAACGTTTG CTGTGTTGAT AACAAATATG CTGGTCCACA TGTCAAACAC
Ugd2 ATTATGCACA TCAACGTTTG CTGTGTTGAT AACAAATATG CTGGTCCACA TGTCAAACAC
Ugd19 ATTATGCACA TCAACGTTTG CTGTGTTGAT AACAAATATG CTGGTCCACA TGTCAAACAC
Ugd16 ATTATTCACA TCAACGTTTG CTGTGTTGAT GACAAACATG CTGGTCCACA TGTCAAACAC
Ugd4 ATTATTCACA TCAACGTTTG CTGTGTTGAT GACAAACATG CTGGTCCACA TGTCAAACAC
Ugd15 ATTATTCACA TCAACGTTCG CTGTGTTGAT GACAAACATG CTGGTCCACA TGTCAAACAC
Ugd23 ATTATTCACA TCAACGTTCG CTGTGTTGAT GACAAACATG CTGGTCCACA TGTCAAACAC
GK18 ATTATTCACA TCAACGTTTG CTGTGTTGAT GACAAACATG CTGGTCCACA TGTCAAACAC
BCBL-R ATTATTCACA TCAACGTTTG CTGTGTTGAT GACAAACATG CTGGTCCACA TGTCAAACAC
Ugd29 ATTATTCACA TCAACGTTTG CTGTGTTGAT GACAAACATG CTGGTCCACA TGTCAAACAC
Cons ATTAT-CACA TCAACGTT-G CTGTGTTGAT -ACAAA-ATG CTGGTCCACA TGTCAAACAC
841 900
Ugd12 CTGGCAATCG TCTTGGATAT TTTTTCCAAT TAGCTTCACT CTCAGCTTGC CATTCTTGTA
Ugd2 CTGGCAATCG TCTTGGATAT TTTTTCCAAT TAGTTTCACT CTCAGCTTGC CATTCTTGTA
Ugd19 CTGGCAATCG TCTTGGATAT TTTTTCCAAT TAGCTTCACT CTCAGCTTGC CATTCTTGTA
Ugd16 CTGGCAATCG TCTTGGATAT TTTTTCCAAT TAGCTTCACT CTCAGCTTGC CATTCTTGTA
Ugd4 CTGGCAATCG TCTTGGATAT TTTTTCCAAT TAGCTTCACT CTCAGCTTGC CATTCTTGTA
Ugd15 CTGGCAATCG TCTTGGATAT TTTTCCCAAT TAGCTTCACT CTCAGCTTGC CATTCTTGTA
Ugd23 CTGGCAATCG TCTTGGATAT TTTTCCCAAT TAGCTTCACT CTCAGCTTGC CATTCTTGTA
GK18 CTGGCAATCG TCTTGGATAT TTTTTCCAAT TAGCTTCACT CTCAGCTTGC CATTCTTGTA
BCBL-R CTGGCAATCG TCTTGGATAT TTTTTCCAAT TAGCTTCACT CTCAGCTTGC CATTCTTGCA
Ugd29 CTGGCAATCG TCTTGGATAT TTTTTCCAAT TAGCTTCACT CTCAGCTTGC CATTCTTGTA
Cons CTGGCAATCG TCTTGGATAT TTTT-CCAAT TAG-TTCACT CTCAGCTTGC CATTCTTG-A


901 960
Ugd12 TGCATTTGCC ACCGTGAAAA CAGGCAATAT TAAATTAGTG TCATCTGTCT CTTTCATTTG
Ugd2 TGCATTTGCC ACCGTGAAAA CAGGCAATAT TAAATTAGTG TCATCTGTCT CTTTCATTTG
Ugd19 TGCATTTGCC ACCGTGAAAA CAGGCAATAT TAAATTAGTG TCATCTGTCT CTTTCATTTG
Ugd16 TGCATTTGCC ACCGTGAAAA CAGGCAATAT TAAATTAGTG TCATCTGTCT CTTTCATTTG
Ugd4 TGCATTTGCC ACCGTGAAAA CAGGCAATAT TAAATTAGTG TCATCTGTCT CTTTCATTTG
Ugd15 TGCATTTGCC ACCGTGAAAA CAGGCAATAT TAAATTAGTG TCATCTGTCT CTTTCATTTG
Ugd23 TGCATTTGCC ACCGTGAAAA CAGGCAATAT TAAATTAGTG TCATCTGTCT CTTTCATTTG
GK18 TGCATTTGCC ACCGTGAAAA CAGGCAATAT TAAATTAGTG TCATCTGTCT CTTTCATTTG
BCBL-R TGCATTTGCC ACCGTGAAAA CAGGCAATAT TAAATTAGTG TCATCTGTCT CTTTCATTTG
Ugd29 TGCATTTGCC ACCGTGAAAA CAGGCAATAT TAAATTAGTG TCATCTGTCT CTTTCATTTG
Cons TGCATTTGCC ACCGTGAAAA CAGGCAATAT TAAATTAGTG TCATCTGTCT CTTTCATTTG
961 1020
Ugd12 TGCAGGTAAG TTAACCACTG TTTTCACACC GGAAATTCAC GGATAGGCCT GTCAACTAAA
Ugd2 TGCAGGTAAG TTAACCACTG TTTTCACACC GGAAATTCAC GGATAGGCCT GTCAACTAAA
Ugd19 TGCAGGTAAG TTAACCACTG TTTTCACACC GGAAATTCAC GGATAGGCCT GTCAACTAAA
Ugd16 TGCAGGTAAG TTAACCACTG TTTTCACACC GGAAATTCAC AGATAGGCCT GTCAACTAAA
Ugd4 TGCAGGTAAG TTAACCACTG TTTTCACACC GGAAATTCAC AGATAGGCCT GTCAACTAAA
Ugd15 TGCAGGTAAG TTAACCACTG TTTTCACACC GGAAATTCAC AGATAGGCCT GTCAACTAAA
Ugd23 TGCAGGTAAG TTAACCACTG TTTTCACACC GGAAATTCAC AGATAGGCCT GTCAACTAAA
GK18 TGCAGGTAAG TTAACCACTG TTTTCACACC GGAAATTCAC AGATAGGCCT GTCAACTAAA
BCBL-R TGCAGGTAAG TTAACCACTG TTTTCACACC GGAAATTCAC AGATAGGCCT GTCAACTAAA
Ugd29 TGCAGGTAAG TTAACCACTG TTTTCACACC GGAAATTCAC AGATAGGCCT GTCAACTAAA
Cons TGCAGGTAAG TTAACCACTG TTTTCACACC GGAAATTCAC -GATAGGCCT GTCAACTAAA
---->
1021 1080
Ugd12 CTGAAGTTAA CACTCGGTTA TCTTTTTTTA TAGGCTTGGT CATGGGTTAC CCTGTATCCT
Ugd2 CTGAAGTTAA CACTCGGTTA TCTTTTTTTA TAGGCTTGGT CATGGGTTAC CCTGTATCCT
Ugd19 CTGAAGTTAA CACTCGGTTA TCTTTTTTTA TAGGCTTGGT CATGGGTTAC CCTGTATCCT
Ugd16 CTGAAGTTAA CACTCGGTTA TCTTTTTTTA TAGGCTTGGT CATGGGTTAC CCTGTATCCT
Ugd4 CTGAAGTTAA CACTCGGTTA TCTTTTTTTA TAGGCTTGGT CATGGGTTAC CCTGTATCCT
Ugd15 CTGAAGTTAA CACTCGGTTA TCTTTTTTTA TAGGCTTGGT CATGGGTTAC CCTGTATCCT
Ugd23 CTGAAGTTAA CACTCGGTTA TCTTTTTTTA TAGGCTTGGT CATGGGTTAC CCTGTATCCT
GK18 CTGAAGTTAA CACTCGGTTA TCTTTTTTTA TAGGCTTGGT CATGGGTTAC CCTGTATCCT
BCBL-R CTGAAGTTAA CACTCGGTTA TC.TTTTTTA TAGGCTTGGT CATGGGTTAC CCTGTATCCT
Ugd29 CTGAAGTTAA CACTCGGTTA TCTTTTTTTA TAGGCTTGGT CATGGGTTAC CCTGTATCCT
Cons CTGAAGTTAA CACTCGGTTA TC-TTTTTTA TAGGCTTGGT CATGGGTTAC CCTGTATCCT
<----------
1081 1140
Ugd12 GCTGTAAAAC GCATACATGT ACTGCCACCG CTGCAGGATT AAGTCTTTCC AGCATTTATT
Ugd2 GCTGTAAAAC GCATACATGT ACTGTCACCG CTGCAGGATT AAGTCTTTCC AGCATTTATT
Ugd19 GCTGTAAAAC GCATACATGT ACTGCCACCG CTGCAGGATT AAGTCTTTCC AGCATTTATT
Ugd16 GCTGTAAAAC GCATACATGT ACTGCCACCG CTGCAGGATT AAGTCTTTCC AGCATTTATT
Ugd4 GCTGTAAAAC GCATACATGT ACTGCCACCG CTGCAGGATT AAGTCTTTCC AGCATTTATT
Ugd15 GCTGTAAAAC GCATACATGT ACTGCCACCG CTGCAGGATT AAGTCTTTCC AGCATTTATT
Ugd23 GCTGTAAAAC GCATACATGT ACTGCCACCG CTGCAGGATT AAGTCTTTCC AGCATTTATT
GK18 GCTGTAAAAC GCATACATGT ACTGCCACCG CTGCAGGATT AAGTCTTTCC AGCATTTATT
BCBL-R GCTGTAAAAC GCATACATGT ACTGCCACCG CTGCAGGATT AAGTCTTTCC AGCATTTATT
Ugd29 GCTGTAAAAC GCATACATGT ACTGCCACCG CTGCAGGATT AAGTCTTTCC AGCATTTATT
Cons GCTGTAAAAC GCATACATGT ACTG-CACCG CTGCAGGATT AAGTCTTTCC AGCATTTATT
1141 1200
Ugd12 TGGGTACAGA ACAAGTATTT TTTAAATACA AATTTTAGTT GGTACATAGT ACATGTTGCT
Ugd2 TGGGTACAGA ACAAGTATTT TTTAAATACA AGTTTTAGTT GGTACATAGT ACATGTTGCT
Ugd19 TGGGTACAGA ACAAGTATTT TTTAAATACA AGTTTTAGTT GGTACATAGT ACATGTTGCT
Ugd16 TGGGTACAGA ACAAGTATTT TTTAAATACA AGTTTTAGTT GGTGCATAGT ACATGTTGCT
Ugd4 TGGGTACAGA ACAAGTATTT TTTAAATACA AGTTTTAGTT GGTGCATAGT ACATGTTGCT
Ugd15 TGGGTACAGA ACAAGTATTT TTTAAATACA AGTTTTAGTT GGTGCATAGT ACATGTTGCT
Ugd23 TGGGTACAGA ACAAGTATTT TTTAAATACA AGTTTTAGTT GGTGCATAGT ACATGTTGCT
GK18 TGGGTACAGA ACAAGTATTT TTTAAATACA AGTTTTAGTT GGTGCATAGT ACATGTTGCT
BCBL-R TGGGTACAGA ACAAGTATTT TTTAAATACA AGTTTTAGTT GGTGCATAGT ACATGTTGCT
Ugd29 TGGGTACAGA ACAAGTATTT TTTAAATACA AGTTTTAGTT GGTGCATAGT ACATGTTGCT
Cons TGGGTACAGA ACAAGTATTT TTTAAATACA A-TTTTAGTT GGT-CATAGT ACATGTTGCT
-->


```
        1201                                                        1260
Ugd12   GTTAACGACA TTTTTTGTAG GATTCACTGG CATAATCAGC ACTCTTCATA AGTCATGGGC
 Ugd2   GTTAACGACA TTTTTTGTAG GATTCACTGG CATAATCAGC ACTCTTCATA AGTCATGGGC
Ugd19   GTTAACGACA TTTTTTGTAG GATTCACTGG CATAATCACC ACTCTTCATA AGTCATGGGC
Ugd16   GTTAACGACA TTTTTTGTAG GATTCACTGG CATAATCAGC ACTCTTCATA AGTCATGGGC
 Ugd4   GTTAACGACA TTTTTTGTAG GATTCACTGG CATAATCAGC ACTCTTCATA AGTCATGGGC
Ugd15   GTTAACGACA TTTTTTGTAG GATTCACTGG CATAATCAGC ACTCTTCATA AGTCATGGGC
Ugd23   GTTAACGACA TTTTTTGTAG GATTCACTGG CATAATCAGC ACTCTTCATA AGTCATGGGC
 GK18   GTTAACGACA TTTTTTGTAG GATTCACTGG CATAATCAGC ACTCTTCATA AGTCATGGGC
BCBL-R  GTTAACGACA TTTTTTGTAG GATTCACTGG CATAATCAGC ACTCTTCATA AGTCATGGGC
Ugd29   GTTAACGACA TTTTTTGTAG GATTCACTGG CATAATCAGC ACTCTTCATA AGTCATGGGC
 Cons   GTTAACGACA TTTTTTGTAG GATTCACTGG CATAATCA-C ACTCTTCATA AGTCATGGGC
                              <-----------------------------------------
        1261                                                        1320
Ugd12   TCCACCAAAA CGTGGAATAT TAACGTTCCT GCTTCTTCAG GGGGGAGTGC TAACCACACA
 Ugd2   TCCACCAAAA CGTGGAATAT TAACGTTCCT GCTTCTTCAG GGGGGAGTGC TAACCACACA
Ugd19   TCCACCAAAA CGTGGAATAT TAACGTTCCT GCTTCTTCAG GGGGGAGTGC TAACCACACA
Ugd16   TCCACCAAAA CGTGGAATAT TAACGTTCCT GCTTCTTCAG GGGGGAGTGC TAACCACACA
 Ugd4   TCCACCAAAA CGTGGAATAT TAACGTTCCT GCTTCTTCAG GGGGGAGTGC TAACCACACA
Ugd15   TCCACCAAAA CGTGGAATAT TAACGTTCCT GCTTCTTCAG GGGGGAGTGC TAACCACACA
Ugd23   TCCACCAAAA CGTGGAATAT TAACGTTCCT GCTTCTTCAG GGGGGAGTGC TAACCACACA
 GK18   TCCACCAAAA CGTGGAATAT TAACGTTCCT GCTTCTTCAG GGGGGAGTGC TAACCACACA
BCBL-R  TCCACCAAAA CGTGGAATAT TAACGTTCCT GCTTCTTCAG GGGGGAGTGC TAACCACACA
Ugd29   TCCACCAAAA CGTGGAATAT TAACGTTCCT GCTTCTTCAG GGGGGAGTGC TAACCACACA
 Cons   TCCACCAAAA CGTGGAATAT TAACGTTCCT GCTTCTTCAG GGGGGAGTGC TAACCACACA
        ------------------------------------------------------------
        1321                                                        1380
Ugd12   AACTTTAACA ACCGAACTAC TTGCAATTAC TAGCACCACT GGCAATATCA AAGGTTTGTT
 Ugd2   AACTTTAACA ACCGAACTAC TTGCAATTAC TAGCACCACT GGCAATATCA AAGGTTTGTT
Ugd19   AACTTTAACA ACCGAACTAC TTGCAATTAC TAGCACCACT GGCAATATCA AAGGTTTGTT
Ugd16   AACTTTAACA ACCGAACTAC TTGCAATTAC TAGCACCACT GGCAATATCA AAGGTTTGTT
 Ugd4   AACTTTAACA ACCGAACTAC TTGCAATTAC TAGCACCACT GGCAATATCA AAGGTTTGTT
Ugd15   AACTTTAACA ACCGAACTAC TTGCAATTAC TAGCACCACT GGCAATATCA AAGGTTTGTT
Ugd23   AACTTTAACA ACCGAACTAC TTGCAATTAC TAGCACCACT GGCAATATCA AAGGTTTGTT
 GK18   AACTTTAACA ACCGAACTAC TTGCAATTAC TAGCACCACT GGCAATATCA AAGGTTTGTT
BCBL-R  AACTTTAACA ACCGAACTAC TTGCAATTAC TAGCACCACT GGCAATATCA AAGGTTTGTT
Ugd29   AACTTTAACA ACCGAACTAC TTGCCATTAC TAGCACCACT GGCAATATCA AAGGTTTGTT
 Cons   AACTTTAACA ACCGAACTAC TTGC-ATTAC TAGCACCACT GGCAATATCA AAGGTTTGTT
        ----------------------------------------------------------->
        1381                                                        1440
Ugd12   GTATAGCATT GATGGTATGG TTCAAGCATC TAAATAAATG GACACTACAT AATTTGTGTT
 Ugd2   GTATAGCATT GATGGTATGG TTCAAGCATC TAAATAAATG GACACTACAT AATTTGTGTT
Ugd19   GTATAGCATT GATGGTATGG TTCAAGCATC TAAATAAATG GACACTACAT AATTTGTGTT
Ugd16   GTATAGCATT GATGGTATGG TTCAAGCATC TAAATAAATG GACACTACAT AATTTGTGTT
 Ugd4   GTATAGCATT GATGGTATGG TTCAAGCATC TAAATAAATG GACACTACAT AATTTGTGTT
Ugd15   GTATAGCATT GATGGTATGG TTCAAGCATC TAAATAAATG GACACTACAT AATTTGTGTT
Ugd23   GTATAGCATT GATGGTATGG TTCAAGCATC TAAATAAATG GACACTACAT AATTTGTGTT
 GK18   GTATAGCATT GATGGTATGG TTCAAGCATC TAAATAAATG TACACTACAT AATTTGTGTT
BCBL-R  GTATAGCATT GATGGTATGG TTCAAGCATC TAAATAAATG GACACTACAT AATTTGTGTT
Ugd29   GTATAGCATT GATGGTATGG TTCAAGCATC TAAATAAATG GACACTACAT AATTTGTGTT
 Cons   GTATAGCATT GATGGTATGG TTCAAGCATC TAAATAAATG -ACACTACAT AATTTGTGTT

        1441                                                        1500
Ugd12   TTTATGTATA AACAGGCCAC GAAATATTGC TTCTTGTATG CCTCATTTTT CTATGGTGCC
 Ugd2   TTTATGTATA AACAGGCCAC GAAATATTGC TTCTTGTATG CCTCATTTTT CTATGGTGCC
Ugd19   TTTATGTATA AACAGGCCAC GAAATATTGC TTCTTGTATG CCTCATTTTT CTATGGTGCC
Ugd16   TTTATGTATA AACAGGCCAC GAAATATTGC TTCTTGTATG CCTCATTTTT CTATGGTGCC
 Ugd4   TTTATGTATA AACAGGCCAC GAAATATTGC TTCTTGTATG CCTCATTTTT CTATGGTGCC
Ugd15   TTTATGTATA AACAGGCCAC GAAATATTGC TTCTTGTATG CCTCATTTTT CTATGGTGCC
Ugd23   TTTATGTATA AACAGGCCAC GAAATATTGC TTCTTGTATG CCTCATTTTT CTATGGTGCC
 GK18   TTTATGTATA AACAGGCCAC GAAATATTGC TTCTTGTATG CCTCATTTTT CTATGGTGCC
BCBL-R  TTTATGTATA AACAGGCCAC GAAATATTGC TTCTTGTATG CCTCATTTTT CTATGGTGCC
Ugd29   TTTATGTATA AACAGGCCAC GAAATATTGC TTCTTGTATG CCTCATTTTT CTATGGTGCC
 Cons   TTTATGTATA AACAGGCCAC GAAATATTGC TTCTTGTATG CCTCATTTTT CTATGGTGCC
                    <-----------------------------------------------
```

```
          1501                                                       1560
Ugd12     TTTATGTGTG GCAAAGTTTT AACAAGACGT CTCTTGTTAC TGGAATACTG CATTTAATTG
Ugd2      TTTATGTGTG GCAAAGTTTT AACAAGGCGT CTCTTGTTAC TGGAATACTG CATTTAATTG
Ugd19     TTTATGTGTG GCAAAGTTTT AACAAGGCGT CTCTTGTTAC TGGAATACTG CATTTAATTG
Ugd16     TTTATGTGTG GCAAAGTTTT AACAAGGCGT CTCTTGTTAC TGGAATGCTG CATTTAATTG
Ugd4      TTTATGTGTG GCAAAGTTTT AACAAGGCGT CTCTTGTTAC TGGAATGCTG CATTTAATTG
Ugd15     TTTATGTGTG GCAAAGTTTT AACAAGGCGT CTCTTGTTAC TGGAATGCTG CATTTAATTG
Ugd23     TTTATGTGTG GCAAAGTTTT AACAAGGCGT CTCTTGTTAC TGGAATGCTG CATTTAATTG
GK18      TTTATGTGTG GCAAAGTTTT AACAAGGCGT CTCTTGTTAC TGGAATGCTG CATTTAATTG
BCBL-R    TTTATGTGTG GCAAAGTTTT AACAAGGCGT CTCTTGTTAC TGGAATGCTG CATTTAATTG
Ugd29     TTTATGTGTG GCAAAGTTTT AACAAGGCGT CTCTTGTTAC TGGAATGCTG CATTTAATTG
Cons      TTTATGTGTG GCAAAGTTTT AACAAG-CGT CTCTTGTTAC TGGAAT-CTG CATTTAATTG
          ------------------------------------------------------------
          1561                                                       1620
Ugd12     CAGGTAGGTA GCACTTTCCA CATTATAAAG TTGGATTTTA TTTTTTAACA ATTACACAAA
Ugd2      CAGGTAGGTA GCACTTTCCA CATTATAAAG TTGGATTTTA TTTTTTAACA ATTACACAAA
Ugd19     CAGGTAGGTA GCACTTTCCA CATTATAAAG TTGGATTTTA TTTTTTAACA ATTACACAAA
Ugd16     CAGGTAGGTA GCACTTTCCA CATTATAAAG TTGGATTTTA TTTTTTAACA ATGACACAAA
Ugd4      CAGGTAGGTA GCACTTTCCA CATTATAAAG TTGGATTTTA TTTTTTAACA ATGACACAAA
Ugd15     CAGGTAGGTA GCACTTTCCA CATTATAAAG TTGGATTTTA TTTTTTAACA ATGACACAAA
Ugd23     CAGGTAGGTA GCACTTTCCA CATTATAAAG TTGGATTTTA TTTTTTAACA ATGACACAAA
GK18      CAGGTAGGTA GCACTTTCCA CATTATAAAG TTGGATTTTA TTTTTTAACA ATGACACAAA
BCBL-R    CAGGTAGGTA GCACTTTCCA CATTATAAAG TTGGATTTTA TTTTTTAACA ATGACACAAA
Ugd29     CAGGTAGGTA GCACTTTCCA CATTATAAAG TTGGATTTTA TTTTTTAACA ATGACACAAA
Cons      CAGGTAGGTA GCACTTTCCA CATTATAAAG TTGGATTTTA TTTTTTAACA AT-ACACAAA
          --->
          1621                                                       1680
Ugd12     AACATAACTA TCCTCTATTT TTTAGCGTGG TCCCACACTG GAGGTTGTGT ACAATTAGTA
Ugd2      AACATAACTA TCCTCTATTT TTTAGCGTGG TCCCACACTG GAGGTTGTGT ACAATTAGTA
Ugd19     AACATAACTA TCCTCTATTT TTTAGCGTGG TCCCACACTG GAGGTTGTGT ACAATTAGTA
Ugd16     AACATAACTA TCCTCTATTT TTTAGCGTGG TCCCACACTG GAGGTTGTGT ACAATTAGTA
Ugd4      AACATAACTA TCCTCTATTT TTTAGCGTGG TCCCACACTG GAGGTTGTGT ACAATTAGTA
Ugd15     AACATAACTA TCCTCTATTT TTTAGCGTGG TCCCACACTG GAGGTTGTGT ACAATTAGTA
Ugd23     AACATAACTA TCCTCTATTT TTTAGCGTGG TCCCACACTG GAGGTTGTGT ACAATTAGTA
GK18      AACATAACTA TCCTCTATTT TTTAGCGTGG TCCCACACTG GAGGTTGTGT ACAATTAGTA
BCBL-R    AACATAACTA TCCTCTATTT TTTAGCGTGG TCCCACACTG GAGGTTGTGT ACAATTAGTA
Ugd29     AACATAACTA TCCTCTATTT TTTAGCGTGG TCCCACACTG GAGGTTGTGT ACAATTAGTA
Cons      AACATAACTA TCCTCTATTT TTTAGCGTGG TCCCACACTG GAGGTTGTGT ACAATTAGTA
                              <-----------------------------------
          1681                                                       1740
Ugd12     ATGCTTCTCC CTAGTGGTTT AACAAGAGGC ATCCTAACAA TGATCATCTG CATCAGTACA
Ugd2      ATGCTTCTCC CTAGTGGTTT AACAAGAGGC ATCCTAACAA TGATCATCTG CATCAGTACA
Ugd19     ATGCTTCTCC CTAGTGGTTT AACAAGAGGC ATCCTAACAA TGATCATCTG CATCAGTACA
Ugd16     ATGCTTCTCC CTAGTGGTTT AACAAGAGGC ATCCTAACAA TGATCATCTG CATCAGTACA
Ugd4      ATGCTTCTCC CTAGTGGTTT AACAAGAGGC ATCCTAACAA TGATCATCTG CATCAGTACA
Ugd15     ATGCTTCTCC CTAGTGGTTT AACAAGAGGC ATCCTAACAA TGATCATCTG CATCAGTACA
Ugd23     ATGCTTCTCC CTAGTGGTTT AACAAGAGGC ATCCTAACAA TGATCATCTG CATCAGTACA
GK18      ATGCTTCTCC CTAGTGGTTT AACAAGAGGC ATCCTAACAA TGATCATCTG CATCAGTACA
BCBL-R    ATGCTTCTCC CTAGTGGTTT AACAAGAGGC ATCCTAACAA TGATCATCTG CATCAGTACA
Ugd29     ATGCTTCTCC CTAGTGGTTT AACAAGAGGC ATCCTAACAA TGATCATCTG CATCAGTACA
Cons      ATGCTTCTCC CTAGTGGTTT AACAAGAGGC ATCCTAACAA TGATCATCTG CATCAGTACA
          ------------------------------------------------------------
          1741                                                       1800
Ugd12     CTATTCAGTA AGTTAAAGAC TAATTACCAT AAATTATATC ACCACATGAC ATATACTATG
Ugd2      CTATTCAGTA AGTTAAAGAC TAATTACCAT AAATTATATC ACCACATGAC ATATACTATG
Ugd19     CTATTCAGTA AGTTAAAGAC TAATTACCAT AAATTATATC ACCACATGAC ATATACTATG
Ugd16     CTATTCAGTA AGTTAAAGAC TAATTACCAT AAATTATATC ACCACATGAC ATATACTATG
Ugd4      CTATTCAGTA AGTTAAAGAC TAATTACCAT AAATTATATC ACCACATGAC ATATACTATG
Ugd15     CTATTCAGTA AGTTAAAGAC TAATTACCAT AAATTATATC ACCACATGAC ATATACTATG
Ugd23     CTATTCAGTA AGTTAAAGAC TAATTACCAT AAATTATATC ACCACATGAC ATATACTATG
GK18      CTATTCAGTA AGTTAAAGAC TAATTACCAT AAATTATATC ACCACATGAC ATATACTATG
BCBL-R    CTATTCAGTA AGTTAAAGAC TAATTACCAT AAATTATATC ACCACATGAC ATATACTATG
Ugd29     CTATTCAGTA AGTTAAAGAC TAATTACCAT AAATTATATC ACCACATGAC ATATACTATG
Cons      CTATTCAGTA AGTTAAAGAC TAATTACCAT AAATTATATC ACCACATGAC ATATACTATG
          ----->
```

```
        1801                                                        1860
Ugd12   TAAAAAAAGT ATTTTTCTAT TTTAAAATTT AGGTACGTTA CAAGGACTGC TGGTATTTTA
 Ugd2   TAAAAAAAGT ATTTTTCTAT TTTAAAATTT AGGTACGTTA CAAGGACTGC TGGTATTTTA
Ugd19   TAAAAAAACT ATTTTTCTAT TTTAAAATTT AGGTACGTTA CAAGGACTGC TGGTATTTTA
Ugd16   TAAAAAAAGT ATTTTTCTAT TTTAAAATTT AGGTACGTTA CAAGGACTGC TGGTATTTTA
 Ugd4   TAAAAAAAGT ATTTTTCTAT TTTAAAATTT AGGTACGTTA CAAGGACTGC TGGTATTTTA
Ugd15   TAAAAAAAGT ATTTTTCTAT TTTAAAATTT AGGTACGTTA CAAGGACTGC TGGTATTTTA
Ugd23   TAAAAAAAGT ATTTTTCTAT TTTAAAATTT AGGTACGTTA CAAGGACTGC TGGTATTTTA
 GK18   TAAAAAAAGT ATTTTTCTAT TTTAAAATTT AGGTACGTTA CAAGGACTGC TGGTATTTTA
BCBL-R  TAAAAAAAGT ATTTTTCTAT TTTAAAATTT AGGTACTTTA CAAGGACTGC TGGTATTTTA
Ugd29   TAAAAAAAGT ATTTTTCTAT TTTAAAATTT AGGTACGTTA CAAGGACTGC TGGTATTTTA
 Cons   TAAAAAAA-T ATTTTTCTAT TTTAAAATTT AGGTAC-TTA CAAGGACTGC TGGTATTTTA
                                          <----------------------------
        1861                                                        1920
Ugd12   TTTATATAAG GAAAAAAAAG TGGTGGCTGT AAATAGTTAC CGACAGAGAA GGCGGCGCAT
 Ugd2   TTTATATAAG GAAAAAAAAG TGGTGGCTGT AAATAGTTAC CGACAGAGAA GGCGGCGCAT
Ugd19   TTTATATAAG GAAAAAAAAG TGGTGGCTGT AAATAGTTAC CGACAGAGAA GGCGGCGCAT
Ugd16   TTTATATAAG GAAAAAACAG TGGTGGCTGT AAATAGTTAC CGACAGAGAA GGCGGCGCAT
 Ugd4   TTTATATAAG GAAAAAAAAG TGGTGGCTGT AAATAGTTAC CGACAGAGAA GGCGGCGCAT
Ugd15   TTTATATAAG GAAAAAAAAG TGGTGGCTGT AAATAGTTAC CGACAGAGAA GGCGGCGCAT
Ugd23   TTTATATAAG GAAAAAAAAG TGGTGGCTGT AAATAGTTAC CGACAGAGAA GGCGGCGCAT
 GK18   TTTATATAAG GAAAAAAAAG TGGTGGCTGT AAATAGTTAC CGACAGAGAA GGCGGCGCAT
BCBL-R  TTTATATAAG GAAAAAAAAG TGGTGGCTGT AAATAGTTAC CGACAGAGAA GGCGGCGCAT
Ugd29   TTTATATAAG GAAAAAAAAG TGTTGGCTGT AAATAGTTAC CGACAGAGAA GGCGGCGCAT
 Cons   TTTATATAAG GAAAAAA-AG TG-TGGCTGT AAATAGTTAC CGACAGAGAA GGCGGCGCAT
        ------------------------------------------------------------------
        1921                                                        1980
Ugd12   ATACACGCGG GACCAGAACT TACACCACAA TGACAACCAC CTTGGCAATA ATGTAATTAG
 Ugd2   ATACACGCGG GACCAGAACT TACACCACAA TGACAACCAC CTTGGCAATA ATGTAATTAG
Ugd19   ATACACGCGG GACCAGAACT TACACCACAA TGACAACCAC CTTGGCAATA ATGTAATTAG
Ugd16   ATACACGCGG GACCAGAACT TACACCACAA TGACAACCAC CTTGGCAATA ATGTAATTAG
 Ugd4   ATACACGCGG GACCAGAACT TACACCACAA TGACAACCAC CTTGGCAATA ATGTAATTAG
Ugd15   ATACACGCGG GACCAGAACT TACACCACAA TGACAACCAC CTTGGCAATA ATGTAATTAG
Ugd23   ATACACGCGG GACCAGAACT TACACCACAA TGACAACCAC CTTGGCAATA ATGTAATTAG
 GK18   ATACACGCGG GACCAGAACT TACACCACAA TGACAACCAC CTTGGCAATA ATGTAATTAG
BCBL-R  ATACACGCGG GACCAGAACT TACACCACAA TGACAACCAC CTTGGCAATA ATGTAATTAG
Ugd29   ATACACGCGG GACCAGAACT TACACCACAA TGACAACCAC CTTGGCAATA ATGTAATTAG
 Cons   ATACACGCGG GACCAGAACT TACACCACAA TGACAACCAC CTTGGCAATA ATGTAATTAG
        ------------------------------------------------------------------
        1981                                                        2040
Ugd12   CCCTCCACCA TTACCACCTT TTTTTAGACA GCCTGTTAGG TTACCTAGCC ATGTCACCGA
 Ugd2   CCCTCCACCA TTACCACCTT TTTTTAGACA GCCTGTTAGG TTACCTAGCC ATGTCACCGA
Ugd19   CCCTCCACCA TTACCACCTT TTTTTAGACA GCCTGTTAGG TTACCTAGCC ATGTCACCGA
Ugd16   CCCTCCACCA TTACCACCTT TTTTTAGACA GCCTGTTAGG TTACCTAGCC ATGTCACCGA
 Ugd4   CCCTCCACCA TTACCACCTT TTTTTAGACA GCCTGTTAGG TTACCTAGCC ATGTCACCGA
Ugd15   CCCTCCACCA TTACCACCTT TTTTTAGACA GCCTGTTAGG TTACCTAGCC ATGTCACCGA
Ugd23   CCCTCCACCA TTACCACCTT TTTTTAGACA GCCTGTTAGG TTACCTAGCC ATGTCACCGA
 GK18   CCCTCCACCA TTACCACCTT TTTTTAGACA GCCTGTTAGG TTACCTAGCC ATGTCACCGA
BCBL-R  CCCTCCACCA TTACCACCTT TTTTTAGACA GCCTGTTAGG TTACCTAGCC ATGTCACCGA
Ugd29   CCCTCCACCA TTACCACCTT TTTTTAGACA GCCTGTTAGG TTACCTAGCC ATGTCACCGA
 Cons   CCCTCCACCA TTACCACCTT TTTTTAGACA GCCTGTTAGG TTACCTAGCC ATGTCACCGA
        ------------------------------------------------------------------
        2041                                                        2100
Ugd12   TAGGGGCCGC GGGAGCCAAC CACTGAACGA GGTGGAGTTA CAGGAAGTGA ATAGGGACCC
 Ugd2   TAGGGGCCGC GGGAGCCAAC CACTGAACGA GGTGGAGTTA CAGGAAGTGA ATAGGGACCC
Ugd19   TAGGGGCCGC GGGAGCCAAC CACTGAACGA GGTGGAGTTA CAGGAAGTGA ATAGGGACCC
Ugd16   TAGGGGCCGC GGGAGCCAAC CACTGAACGA GGTGGAGTTA CAGGAAGTGA ATAGGGACCC
 Ugd4   TAGGGGCCGC GGGAGCCAAC CACTGAACGA GGTGGAGTTA CAGGAAGTGA ATAGGGACCC
Ugd15   TAGGGGCCGC GGGAGCCAAC CACTGAACGA GGTGGAGTTA CAGGAAGTGA ATAGGGACCC
Ugd23   TAGGGGCCGC GGGAGCCAAC CACTGAACGA GGTGGAGTTA CAGGAAGTGA ATAGGGACCC
 GK18   TAGGGGCCGC GGGAGCCAAC TACTGAACGA GGTGGAGTTA CAGGAAGTGA ATAGGGACCC
BCBL-R  TAGGGGCCGC GGGAGCCAAC CACTGAACGA GGTGGAGTTA CAGGAAGTGA ATAGGGACCC
Ugd29   TAGGGGCCGC GGGAGCCAAC CACTGAACGA GGTGGAGTTA CAGGAAGTGA ATAGGGACCC
 Cons   TAGGGGCCGC GGGAGCCAAC -ACTGAACGA GGTGGAGTTA CAGGAAGTGA ATAGGGACCC
        ------------------------------------------------------------------
```

```
        2101                                                        2160
Ugd12   ACCCAATGTA TTCGGGTATG CCAGTATTTT AGTGTCCGGT GCGGAAGAAT CACGTGAACC
 Ugd2   ACCCAATGTA TTCGGGTATG CCAGTATTTT AGTGTCCGGT GCGGAAGAAT CACGTGAACC
Ugd19   ACCCAATGTA TTCGGGTATG CCAGTATTTT AGTGTCCGGT GCGGAAGAAT CACGTGAACC
Ugd16   ACCCAATGTA TTCGGGTATG CCAGTATTTT AGTGTCCGGT GCGGAAGAAT CACGTGAACC
 Ugd4   ACCCAATGTA TTCGGGTATG CCAGTATTTT AGTGTCCGGT GCGGAAGAAT CACGTGAACC
Ugd15   ACCCAATGTA TTCGGGTATG CCAGTATTTT AGTGTCCGGT GCGGAAGAAT CACGTGAACC
Ugd23   ACCCAATGTA TTCGGGTATG CCAGTATTTT AGTGTCCGGT GCGGAAGAAT CACGTGAACC
 GK18   ACCCAATGTA TTCGGGTATG CCAGTATTTT AGTGTCCGGT GCGGAAGAAT CACGTGAACC
BCBL-R  ACCCAATGTA TTCGGGTATG CCAGTATTTT AGTGTCCGGT GCGGAAGAAT CACGTGAACC
Ugd29   ACCCAATGTA TTCGGGTATG CCAGTATTTT AGTGTCCGGT GCGGAAGAAT CACGTGAACC
 Cons   ACCCAATGTA TTCGGGTATG CCAGTATTTT AGTGTCCGGT GCGGAAGAAT CACGTGAACC
        ------------------------------------------------------------
        2161                                                        2220
Ugd12   ATCACCACAG CCAGACCAAT CAGGAATGTC AATTTTAAGG GTTGATGGTG GGAGTGCCTT
 Ugd2   ATCACCACAG CCAGACCAAT CAGGAATGTC AATTTTAAGG GTTGATGGTG GGAGTGCCTT
Ugd19   ATCACCACAG CCAGACCAAT CAGGAATGTC AATTTTAAGG GTTGATGGTG GGAGTGCCTT
Ugd16   ATCACCACAG CCAGACCAAT CAGGAATGTC AATTTTAAGG GTTGATGGTG GGAGTGCCTT
 Ugd4   ATCACCACAG CCAGACCAAT CAGGAATGTC AATTTTAAGG GTTGATGGTG GGAGTGCCTT
Ugd15   ATCACCACAG CCAGACCAAT CAGGAATGTC AATTTTAAGG GTTGATGGTG GGAGTGCCTT
Ugd23   ATCACCACAG CCAGACCAAT CAGGAATGTC AATTTTAAGG GTTGATGGTG GGAGTGCCTT
 GK18   ATCACCACAG CCAGACCAAT CAGGAATGTC AATTTTAAGG GTTGATGGTG GGAGTGCCTT
BCBL-R  ATCACCACAG CCAGACCAAT CAGGAATGTC AATTTTAAGG GTTGATGGTG GGAGTGCCTT
Ugd29   ATCACCACAG CCAGACCAAT CAGGAATGTC AATTTTAAGG GTTGATGGTG GGAGTGCCTT
 Cons   ATCACCACAG CCAGACCAAT CAGGAATGTC AATTTTAAGG GTTGATGGTG GGAGTGCCTT
        ------------------------------------------------------------
        2221                                                        2280
Ugd12   CCGTATAGAC ACCGCCCAAG CCGCCACCCA ACCGACAGAC GACCTGTATG AGGAGGTTTT
 Ugd2   CCGTATAGAC ACCGCCCAAG CCGCCACCCA ACCGACAGAC GACCTGTATG AGGAGGTTTT
Ugd19   CCGTATAGAC ACCGCCCAAG CCGCCACCCA ACCGACAGAC GACCTGTATG AGGAGGTTTT
Ugd16   CCGTATAGAC ACTGCCCAAG CCGCCACCCA ACCGACAGAC GACCTGTATG AGGAGGTTTT
 Ugd4   CCGTATAGAC ACCGCCCAAG CCGCCACCCA ACCGACAGAC GACCTGTATG AGGAGGTTTT
Ugd15   CCGTATAGAC ACCGCCCAAG CCGCCACCCA ACCGACAGAC GACCTGTATG AGGAGGTTTT
Ugd23   CCGTATAGAC ACCGCCCAAG CCGCCACCCA ACCGACAGAC GACCTGTATG AGGAGGTTTT
 GK18   CCGTATAGAC ACCGCCCAAG CCGCCACCCA ACCGACAGAC GACCTGTATG AGGAGGTTTT
BCBL-R  CCGTATAGAC ACCGCCCAAG CCGCCACCCA ACCGACAGAC GACCTGTATG AGGAGGTTTT
Ugd29   CCGTATAGAC ACCGCCCAAG CCGCCACCCA ACCGACAGAC GACCTGTATG AGGAGGTTTT
 Cons   CCGTATAGAC AC-GCCCAAG CCGCCACCCA ACCGACAGAC GACCTGTATG AGGAGGTTTT
        ------------------------------------------------------------
        2281          stop                                          2340
Ugd12   ATTTCCCAGG AACTAGCCCT CCACGACCAC AGACTTTTTG ACATCGATAC CTTTTTTTGA
 Ugd2   ATTTCCCAGG AACTAGCCCT CCACGACCAC AGACTTTTTG ACATCGATAC CTTTTTTTGA
Ugd19   ATTTCCCAGG AACTAGCCCT CCACGACCAC AGACTTTTTG ACATCGATAC CTTTTTTTGA
Ugd16   ATTTCCCAGG AACTAGCCCT CCACGACCAC AGACTTTTTG ACATCGATAC CTTTTTTTGA
 Ugd4   ATTTCCCAGG AACTAGCCCT CCACGACCAC AGACTTTTTG ACATCGATAC CTTTTTTTGA
Ugd15   ATTTCCCAGG AACTAGCCCT CCACGACCAC AGACTTTTTG ACATCGATAC CTTTTTTTGA
Ugd23   ATTTCCCAGG AACTAGCCCT CCACGACCAC AGACTTTTTG ACATCGATAC CTTTTTTTGA
 GK18   ATTTCCCAGG AACTAGCCCT CCACGACCAC AGACTTTTTG ACATCGATAC CTTTTTTTGA
BCBL-R  ATTTCCCAGG AACTAGCCCT CCACGACCAC AGACTTTTTG ACATCGATAC CTTTTTTTGA
Ugd29   ATTTCCCAGG AACTAGCCCT CCACGACCAC AGACTTTTTG ACATCGATAC CTTTTTTTGA
 Cons   ATTTCCCAGG AACTAGCCCT CCACGACCAC AGACTTTTTG ACATCGATAC CTTTTTTTGA
        ----------------><-----------25 bp---------->
        2341                                                        2400
Ugd12   GTATTTGAGG TTAGTGACAT GGCTACATGT AACTGTGGAT TCCACGAAAG CGAAAACAAA
 Ugd2   GTATTTGAGG TTAGTGACAT GGCTACATGT AACTGTGGAT TCCACGAAAG CGAAAACAAA
Ugd19   GTATTTGAGG TTAGTGACAT GGCTACATGT AACTGTGGAT TCCACGAAAG CGAAAACAAA
Ugd16   GTATTTGAGG TTAGTGACAT GGCTACATGT AACTGTGGAT TCGACGAAAG CGAAAACAAA
 Ugd4   GTATTTGAGG TTAGTGACAT GGCTACATGT AACTGTGGAT TCGACGAAAG CGAAAACAAA
Ugd15   GTATTTGAGG TTAGTGACAT GGCTACATGT AACTGTGGAT TCCACGAAAG CGAAAACAAA
Ugd23   GTATTTGAGG TTAGTGACAT GGCTACATGT AACTGTGGAT TCCACGAAAG CGAAAACAAA
 GK18   GTATTTGAGG TTAGTGACAT GGCTACATGT AACTGTGGAT TCCACGAAAG CGAAAACAAA
BCBL-R  GTATTTGAGG TTAGTGACAT GGCTACATGT AACTGTGGAT TCCACGAAAG CGAAAACAAA
Ugd29   GTATTTGAGG TTAGTGACAT GGCTACATGT AACTGTGGAT TCCACGAAAG CGAAAACAAA
 Cons   GTATTTGAGG TTAGTGACAT GGCTACATGT AACTGTGGAT TC-ACGAAAG CGAAAACAAA
```

```
        2401                                                        2460
 Ugd12  CTCTGCAGAG CAACCTGTGC CTGCGTAATA TTTGGGGATT CTCAGTTTCT TCCTACTCCC
  Ugd2  CTCTGCAGAG CAACCTGTGC CTGCGTAATA TTTGGGGATT CTCAGTTTCT TCCTACTCCC
 Ugd19  CTCTACAGAG CAACCTGTGC CTGCGTAATA TTTGGGGATT CTCAGTTTCT TCCTACTCCC
 Ugd16  CTCTGCAGAG CAACCTGTGC CTGCGTAATA TTTGGGGATT CTCAGTTTCT TCCTACTCCC
  Ugd4  CTCTGCAGAG CAACCTGTGC CTGCGTAATA TTTGGGGATT CTCAGTTTCT TCCTACTCCC
 Ugd15  CTCTGCAGAG CAACCTGTGC CTGCGTAATA TTTGGGGATT CTCAGTTTCT TCCTACTCCC
 Ugd23  CTCTGCAGAG CAACCTGTGC CTGCGTAATA TTTGGGGATT CTCAGTTTCT TCCTACTCCC
  GK18  CTCTGCAGAG CAACCTGTGC CTGCGTAATA TTTGGGGATT CTCAGTTTCT TCCTACTCCC
BCBL-R  CTCTGCAGAG CAACCTGTGC CTGCGTAATA TTTGGGGATT CTCAGTTTCT TCCTACTCCC
 Ugd29  CTCTGCAGAG CAACCTGTGC CTGCGTAATA TTTGGGGATT CTCAGTTTCT TCCTACTCCC
  Cons  CTCT-CAGAG CAACCTGTGC CTGCGTAATA TTTGGGGATT CTCAGTTTCT TCCTACTCCC

        2461                                                 2505
 Ugd12  CAGAGAAATC AAAGCCCTAA CCCAAGTCTG ACTACAGAGG GTGTC
  Ugd2  CAGAGAAATC AAAGCCCTAA CCCAAGTCTG ACTACAGAGG GTGTC
 Ugd19  CAGAGAAATC AAAGCCCTAA CCCAAGTCTG ACTACAGAGG GTGTC
 Ugd16  CAGAGAAATC AAAGCCCTAA CCCAAGTCTG ACTACAGAGG GTGTC
  Ugd4  CAGAGAAATC AAAGCCCTAA CCCAAGTCTG ACTACAGAGG GTGTC
 Ugd15  CAGAGAAATC AAAGCCCTAA CCCAAGTCTG ACTACAGAGG GTGTC
 Ugd23  CAGAGAAATC AAAGCCCTAA CCCAAGTCTG ACTACAGAGG GTGTC
  GK18  CAGAGAAATC AAAGCCCTAA CCCAAGTCTG ACTACAGAGG GTGTC
BCBL-R  CAGAGAAATC AAAGCCCTAA CCCAAGTCTG ACTACAGAGG GTGTC
 Ugd29  CAGAGAAATC AAAGCCCTAA CCCAAGTCTG ACTACAGAGG GTGTC
  Cons  CAGAGAAATC AAAGCCCTAA CCCAAGTCTG ACTACAGAGG GTGTC
```

divergence from the M allele commencing 25 bp downstream from the K15 stop codon (position 2321 in Fig. 5.5). Sites of nucleotide and amino acid changes (in K15 alone) are shown in Fig. 5.4B.

A total of 34 substitutions (Fig. 5.4B) were identified in the gene collection in addition to a 3-bp deletion (nt 226-228; Fig. 5.5) in the first exon of Ugd29 and a 10-bp insertion (nt 433-442; Fig. 5.5) in the first intron of Ugd2, Ugd12 and Ugd19. In coding regions, nine substitutions are synonymous and 16 are non-synonymous; nine substitutions are within introns (Fig. 5.4B). A cluster of four sites plus the 3 bp deletion in Ugd29 are present within the first six codons (Fig. 5.5). Two substitutions are located in the downstream flanking region and none in the upstream region (Fig. 5.5).

Overall, the nucleotide changes divide the P strains into two distinct groups: Ugd2, Ugd12 and Ugd19; and Ugd4, Ugd15, Ugd16, Ugd23, Ugd29, BCBL-R and GK18. All ten sequences (including BCBL-R and GK18) are different from each other (see below chapter 6; Fig. 6.3D). Pairwise sequence comparisons (shown Table 5.1) show an overall nucleotide divergence among the sequences ranging from 0.05% (between Ugd15 and Ugd23) to 0.9% (between BCBL-R/GK18/Ugd16 and Ugd2/Ugd12/Ugd19). Divergence within each of the two main groups (Ugd2/12/19 and the other group) is minimal, 0.2% and 0.05-0.4%, respectively. The divergence between the two groups is 0.8-0.9%. Thus, the most widely diverged P sequences differ by 0.9%.

5.4.3 Conclusion

The results indicate that the sequences of the M allele in Ugd10 and BC-1 differ by a value (1.2%) that is similar to that (0.9%) for the two most widely diverged representatives of the P allele for which sequence data are available.

Table 5.1. Percent pairwise divergences among the K15 P sequences

Sequence	BCBL-R	GK18	Ugd4	Ugd16	Ugd15	Ugd23	Ugd29	Ugd2	Ugd12	Ugd19
BCBL-R	0	0.2	0.2	0.3	0.2	0.3	0.2	0.9	0.9	0.9
GK18		0	0.3	0.4	0.3	0.3	0.3	0.9	0.9	0.9
Ugd4			0	0.1	0.2	0.2	0.2	0.8	0.8	0.8
Ugd16				0	0.3	0.3	0.3	0.9	0.9	0.9
Ugd15					0	0.05	0.2	0.8	0.8	0.8
Ugd23						0	0.2	0.9	0.9	0.9
Ugd29							0	0.8	0.8	0.8
Ugd2								0	0.2	0.2
Ugd12									0	0.2
Ugd19										0

5.5 CONCLUSION AND DISCUSSION

5.5.1 K15 genotypes in Uganda

The majority of HHV-8-positive samples from Uganda contains the P allele, as was predicted by Poole et al. (1999) from data on Eastern Africa samples. This allele is associated with all the K1 subtypes (A5, B and C) identified in the Ugandan samples in this study. The M allele was identified in a single sample (Ugd10, which has a K1 B subtype) originating from a 70-yr old male patient with endemic KS from Western Uganda, a region where endemic KS was common in older men prior to the HIV epidemic (Ziegler and Katongole-Mbidde, 1996). To my knowledge, this is the first time the K15 M allele has been reported in Eastern Africa samples. Therefore, this allele appears to be rarer in Eastern Africa than in Central and West Africa as found by Lacoste et al. (2000a).

Two samples, Ugd5 and Ugd14, gave no product with either M- or P-specific primers, suggesting that they might contain a non-P, non-M allele, or that they are deleted in the region amplified. However, it is also possible that sub-optimal PCR conditions were responsible for this result, e.g. sub-optimal levels of the template resulting from low DNA concentration (as may be the case with Ugd5) or low HHV-8 DNA copy numbers (as may be the case with Ugd14).

5.5.2 Divergence within the K15 gene

The data on the sequences of the M and P alleles indicate that each of these alleles has been evolving in at least two forms. The level of divergence (1.2%) detected in the sequences with the M allele (Ugd10 and BC-1) was surprising in view of the reports by Poole et al. (1999) and Hayward (1999). They claimed

that the M sequences in the strains they analysed were virtually identical to that in BC-1. This observation also rules out the possibility that the variation in the BC-1 sequence is an artifact of adaptation to growth in vitro. The maximum divergence within the P allele sequences (0.9%) is similar to that between the two M allele sequences.

CHAPTER 6
RESULTS AND DISCUSSION-IV
CHARACTERIZATION OF OTHER LOCI

6.1 INTRODUCTION

Poole et al. (1999) analysed sequence variation of more than 60 strains at gene loci ORF26, T0.7/K12 and ORF75. Subtypes were defined based on nucleotide patterns. Sequence variation at these loci was much less than that in K1, but there was a general correlation between the groups defined at these loci and the K1 groups. Nevertheless, sequences of some strains belonging to certain K1 subtypes could not be distinguished. In ORF26, some K1 B strains had sequences indistinguishable from those of K1 C strains (a pattern designated B/C), also K1 D strains had sequences indistinguishable from those of K1 subtype A strains (D/A pattern). Similarly, some K1 A strains had sequences indistinguishable from those of K1 C strains at T0.7/K12 (A/C), while K1 subtype A and C strains had indistinguishable ORF75 sequences (also referred to as A/C). Furthermore, in ORF75, eight of 14 K1 B strains analysed also had the A/C nucleotide pattern. In genomes with the K15 M allele, Poole et al. (1999) noted a nucleotide pattern (designated M) that was associated with the presence of the K15 M allele and extended into ORF75 and, in some cases, T0.7/K12.

Alagiozoglou et al. (2000) analysed the phylogenetic relationships of HHV-8 strains in South Africa using the ORF75 gene. These investigators found that, in addition to the previously identified A (later renamed A/C by Poole et al., 1999), B and C (also later renamed M by Poole et al., 1999) subtypes, a novel subtype (designated N) is circulating in South Africa. The old nomenclature for ORF75 groups used by Alagiozoglou et al. (2000) was derived from previous work by

Zong et al. (1997). The revised nomenclature (Poole et al., 1999) is used in this study.

Evidence for recombination was noted in 20-30% of the genomes studied by Poole et al. (1999). Approximately half (five of 12) of African strains displayed mosaic genomes that probably reflect a history of recombination within and among subtypes. All five strains (including ST1 from Uganda) had the A/C pattern at ORF75. Furthermore, five of the 12 African strains (including ST1) had the B/C pattern in ORF26. ST2 (the only other Ugandan strain in this study) had subtype B patterns in ORF26 and ORF75, and this genome was interpreted as being subtype B throughout. However, the patterns of ST1 and ST2 were not determined in T0.7/K12 (and K15, as seen in section 5.1 above). A strain from Tanzania (OKS3; Table 2.1) with an A5 K1 gene had what was described as being a mosaic genome, with a subtype B pattern in ORF26 and T0.7/K12, an A/C pattern in ORF75 and a K15 P allele. On the other hand, 431KAP (Table 2.1) from Zaire had subtype B patterns at K1, ORF26, T0.7/K12 and ORF75 and was linked to a K15 P allele. This strain was described as being subtype B throughout.

Evidence for recombination has also been observed in EBV. Three intertypic recombinants with type 1 sequences at the EBNA2 locus and type 2 sequences at some or all of the EBNA-3A, -3B, -3C loci were identified among 34 EBV isolates from non-immunocompromised Chinese donors (Midgley et al., 2000).

Apart from the two Ugandan HHV-8 strains (ST1 and ST2) whose genomes have been characterised at two other gene loci (in addition to K1), no other Ugandan strain has been characterised at multiple loci. The objective of this study was to analyse sequence variation in strains from Uganda at five additional loci (K3,

ORF26, K9, T0.7/K12 and ORF75). The overall genotypes of these strains were then evaluated, shedding further light on the evolution of HHV-8.

6.2 PCR AMPLIFICATION

PCR products were generated from loci K3, ORF26, K9, T0.7/K12 and ORF75 of the nine strains (Ugd2, Ugd4, Ugd10, Ugd12, Ugd15, Ugd16, Ugd19, Ugd23 and Ugd29) whose K15 genes had been sequenced. The loci selected were spaced out at roughly equal intervals across the genome (K3 at 0.14 map units (MU), ORF26 at 0.35 MU, K9 at 0.61 MU, T0.7/K12 at 0.85 MU and ORF75 at 0.97MU), and also included loci examined in previous studies (Alagiozoglou et al., 2000; Poole et al., 1999). The coordinates are given in Table 2.2. The portion of ORF75 immediately adjacent to K15 was chosen for analysis with the view that this might shed more light on the evolution of the K15 gene region. For ORF26, PCR products were also generated from all the other Ugandan samples. PCR products were also generated for K9, T0.7/K12 and ORF75 loci of Ugd30. The primers used and sizes of the products are shown in Table 2.3. ORF26 products used in this study were those amplified by KS4/KS5.

Representative PCR products are shown in Fig. 6.1. All the loci gave strong signals for all samples, and this allowed direct sequencing of the PCR products. The segments consisted entirely of coding sequences, except for T0.7/K12, which consisted of a portion of the K12 gene (aa 13-60 [end]) plus the non-coding sequences downstream from the K12 stop codon.

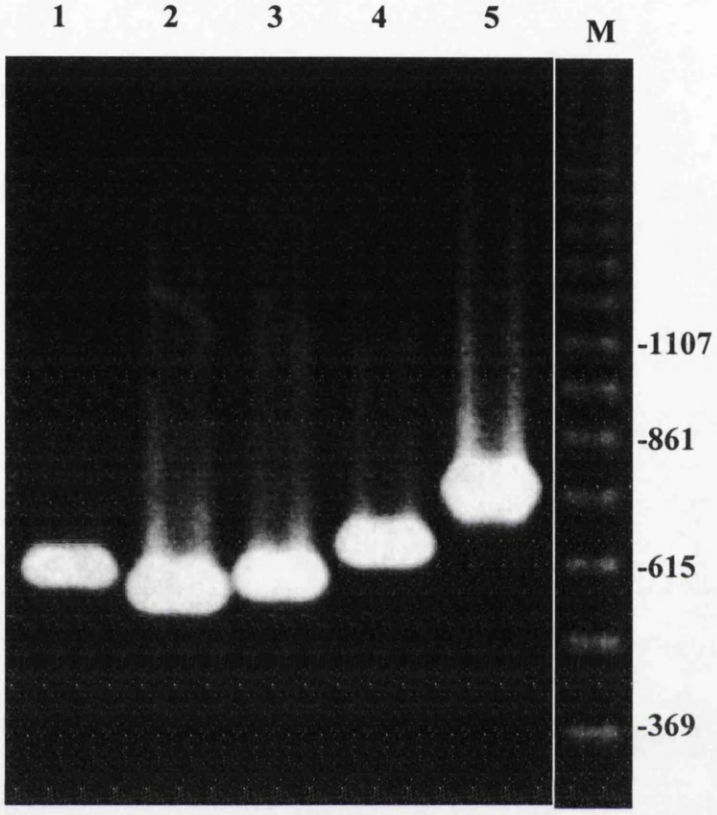


Fig. 6.1. K3, ORF26, K9, T0.7/K12, ORF75 PCR products.

EtBr-stained 1% (w/v) agarose gels showing PCR products of K3 (635 bp), ORF26 (571 bp), K9 (594 bp), T0.7/K12 (648 bp), ORF75 (749 bp) in lanes 1-5, respectively. The products were amplified using primers shown in Table 2.3. M, 123 bp DNA ladder.

6.3 NETWORK ANALYSIS

6.3.1 Genotypes at each loci

The Ugandan sequences were aligned together with reference strains, BC-1 (for all loci), and BCBL-R and GK18 (for ORF75). The alignments are shown in Fig. 6.2. Networks were then drawn to determine the relationships of the strains at each locus. Those for K3 and ORF26 were similar, and therefore were combined (Fig. 6.3A). That for K9 (with only one substitution) was combined with that for T0.7/K12 (Fig. 6.3B). The K15 M allele group (Ugd10 and BC-1) represents a third group not shown at the K15 locus because it is very distant from the P allele.

The groups defined for each locus (together with those for K1) are shown in Table 6.1. Ugd2, Ugd12 and Ugd19 co-segregate at all loci (excluding K1), as do Ugd4, Ugd15, Ugd16, Ugd23, Ugd29 (with the exception of Ugd23 at K3-ORF26 locus). Ugd10 groups with Ugd4/15/16/29 at locus K3-ORF26, but shows links to BC-1 in K9-T0.7/K12, ORF75 and K15.

6.3.2 Evidence for recombination in Ugandan strains

Evaluation of the groups defined by the network analysis together with the phylogenetic groups in K1 indicates that some strains group differently at different loci. Firstly, in K1, the three strains (Ugd4, Ugd12, and Ugd16) that have A5 K1 genes are closer to BC-1 than to other Ugandan strains with B K1 genes. These, however, cluster with other Ugandan strains, and not with BC-1, in K3-ORF26, K9-T0.7/K12, ORF75 and K15, providing evidence for recombination between K1 and K3. Secondly, Ugd10, which does not cluster with BC-1 in K1 and K3-ORF26, appears to be linked to this strain in K9-

Fig. 6.2.Alignments of (A) K3, (B) ORF26, (C) K9, (D) T0.7/K12 and (E) ORF75 DNA sequences of Ugandan and reference strains.

The alignments are in reverse orientation to the genomic sequence except for K3 and ORF26. The coordinates for each loci are given in Table 2.2. Positions of nucleotide changes are shown in bold-red, with the deletion in D represented by dots. Locations of mutated codons are highlighted by open boxes. Asterisks denote synonymous substitutions and amino acid changes are shown for non-synonymous substitutions. The start (E) and stop (D) codons are highlighted in bold-blue.

A - K3

```
                                              *            T/A     60
       1
Ugd10  ACGGGGAAGA ATTTCCTGAA GCTCGATCTC CTCTACCGCA CACTCTGGTG ATGCCGGCCG
Ugd4   ACGGGGAAGA ATTTCCTGAA GCTCGATCTC CTCTACCGCA CACTCTGGTG ATGCCGGCCG
Ugd16  ACGGGGAAGA ATTTCCTGAA GCTCGATCTC CTCTACCGCA CACTCTGGTG ATGTCGGCCG
Ugd29  ACGGGGAAGA ATTTCCTGAA GCTCGATCTC CTCTACCGCA CACTCTGGTG ATGTCGGCCG
Ugd12  ACGGGGAAGA ATTTCCTGAA GCTCGATCTC CTCTACCGCA CACTCTGGTG ATGTCGGCCG
Ugd19  ACGGGGAAGA ATTTCCTGAA GCTCGATCTC CTCTACCGCA CACTCTGGTG ATGTCGGCCG
Ugd2   ACGGGGAAGA ATTTCCTGAA GCTCGATCTC CTCTACCGCA CACTCTGGTG ATGTCGGCCG
Ugd15  ACGGGGAAGA ATTTCCTGAA GCTCGATCTC CTCTACCGCA CACTCTGGTG ATGTCGGCCG
Ugd23  ACGGGGAAGA ATTTCCTGAA GCTCGATCTC CTCTACCGCA CACTCTGGTG ATGTCGGCCG
BC-1   ACGGGGAAGA ATTTCCTGAA GCTCGATCTC CTCTACTGCA CACTCTGGTG ATGTCGGCCG
Cons   ACGGGGAAGA ATTTCCTGAA GCTCGATCTC CTCTAC-GCA CACTCTGGTG ATG-CGGCCG

       61                                     *                   120
Ugd10  AGGTCTATAT GGAAACACTT CAACCCGCGT GTTTACAGCA GCGTATGCCC GCCCCACGTG
Ugd4   AGGTCTATAT GGAAACACTT CAACCCGCGT GTTTACAGCA GCGTATGCCC GCCCCACGTG
Ugd16  AGGTCTATAT GGAAACACTT CAACCCGCGT GTTTACAGCA GCGTATGCCC GCCCCACGTG
Ugd29  AGGTCTATAT GGAAACACTT CAACCCGCGT GTTTACAGCA GCGTATGCCC GCCCCACGTG
Ugd12  AGGTCTATAT GGAAACACTT CAACCCGCGT GTTTACTGCA GCGTATGCCC GCCCCACGTG
Ugd19  AGGTCTATAT GGAAACACTT CAACCCGCGT GTTTACTGCA GCGTATGCCC GCCCCACGTG
Ugd2   AGGTCTATAT GGAAACACTT CAACCCGCGT GTTTACTGCA GCGTATGCCC GCCCCACGTG
Ugd15  AGGTCTATAT GGAAACACTT CAACCCGCGT GTTTACAGCA GCGTATGCCC GCCCCACGTG
Ugd23  AGGTCTATAT GGAAACACTT CAACCCGCGT GTTTACAGCA GCGTATGCCC GCCCCACGTG
BC-1   AGGTCTATAT GGAAACACTT CAACCCGCGT GTTTACAGCA GCGTATGCCC GCCCCACGTG
Cons   AGGTCTATAT GGAAACACTT CAACCCGCGT GTTTAC-GCA GCGTATGCCC GCCCCACGTG

       121                                                        180
Ugd10  GCGCATCATG TGGAAAAACG CACCCAACCC AAAAACGACA AACAATTGGT AAAACACGAA
Ugd4   GCGCATCATG TGGAAAAACG CACCCAACCC AAAAACGACA AACAATTGGT AAAACACGAA
Ugd16  GCGCATCATG TGGAAAAACG CACCCAACCC AAAAACGACA AACAATTGGT AAAACACGAA
Ugd29  GCGCATCATG TGGAAAAACG CACCCAACCC AAAAACGACA AACAATTGGT AAAACACGAA
Ugd12  GCGCATCATG TGGAAAAACG CACCCAACCC AAAAACGACA AACAATTGGT AAAACACGAA
Ugd19  GCGCATCATG TGGAAAAACG CACCCAACCC AAAAACGACA AACAATTGGT AAAACACGAA
Ugd2   GCGCATCATG TGGAAAAACG CACCCAACCC AAAAACGACA AACAATTGGT AAAACACGAA
Ugd15  GCGCATCATG TGGAAAAACG CACCCAACCC AAAAACGACA AACAATTGGT AAAACACGAA
Ugd23  GCGCATCATG TGGAAAAACG CACCCAACCC AAAAACGACA AACAATTGGT AAAACACGAA
BC-1   GCGCATCATG TGGAAAAACG CACCCAACCC AAAAACGACA AACAATTGGT AAAACACGAA
Cons   GCGCATCATG TGGAAAAACG CACCCAACCC AAAAACGACA AACAATTGGT AAAACACGAA

       181                                                        240
Ugd10  AAAAACGTAG TACGCGGCTG CAGCGACGTG ATCTATCTCT GGGTCATGAC CGCCCACTAT
Ugd4   AAAAACGTAG TACGCGGCTG CAGCGACGTG ATCTATCTCT GGGTCATGAC CGCCCACTAT
Ugd16  AAAAACGTAG TACGCGGCTG CAGCGACGTG ATCTATCTCT GGGTCATGAC CGCCCACTAT
Ugd29  AAAAACGTAG TACGCGGCTG CAGCGACGTG ATCTATCTCT GGGTCATGAC CGCCCACTAT
Ugd12  AAAAACGTAG TACGCGGCTG CAGCGACGTG ATCTATCTCT GGGTCATGAC CGCCCACTAT
Ugd19  AAAAACGTAG TACGCGGCTG CAGCGACGTG ATCTATCTCT GGGTCATGAC CGCCCACTAT
Ugd2   AAAAACGTAG TACGCGGCTG CAGCGACGTG ATCTATCTCT GGGTCATGAC CGCCCACTAT
Ugd15  AAAAACGTAG TACGCGGCTG CAGCGACGTG ATCTATCTCT GGGTCATGAC CGCCCACTAT
Ugd23  AAAAACGTAG TACGCGGCTG CAGCGACGTG ATCTATCTCT GGGTCATGAC CGCCCACTAT
BC-1   AAAAACGTAG TACGCGGCTG CAGCGACGTG ATCTATCTCT GGGTCATGAC CGCCCACTAT
Cons   AAAAACGTAG TACGCGGCTG CAGCGACGTG ATCTATCTCT GGGTCATGAC CGCCCACTAT

       241                                                 *      300
Ugd10  ATATAGCCAA ACCCACGTCG CAGCGGCAAG GCCAGCGGCC CCCAATGTCA TGATGAAAAT
Ugd4   ATATAGCCAA ACCCACGTCG CAGCGGCAAG GCCAGCGGCC CCCAATGTCA TGATGAAAAT
Ugd16  ATATAGCCAA ACCCACGTCG CAGCGGCAAG GCCAGCGGCC CCCAATGTCA TGATGAAAAT
Ugd29  ATATAGCCAA ACCCACGTCG CAGCGGCAAG GCCAGCGGCC CCCAATGTCA TGATGAAAAT
Ugd12  ATATAGCCAA ACCCACGTCG CAGCGGCAAG GCCAGCGGCC CCCAATGTCA TGATGAAAAT
Ugd19  ATATAGCCAA ACCCACGTCG CAGCGGCAAG GCCAGCGGCC CCCAATGTCA TGATGAAAAT
Ugd2   ATATAGCCAA ACCCACGTCG CAGCGGCAAG GCCAGCGGCC CCCAATGTCA TGATGAAAAT
Ugd15  ATATAGCCAA ACCCACGTCG CAGCGGCAAG GCCAGCGGCC CCCAATGTCA TGATGAAAAT
Ugd23  ATATAGCCAA ACCCACGTCG CAGCGGCAAG GCCAGCGGCC CCCAATGTCA TGATGAAAAT
BC-1   ATATAGCCAA ACCCACGTCG CAGCGGCAAG GCCAGCGGCC CCCAATGTCA TAATGAAAAT
Cons   ATATAGCCAA ACCCACGTCG CAGCGGCAAG GCCAGCGGCC CCCAATGTCA T-ATGAAAAT
```

```
        301          *                          *                    *          360
Ugd10   AAAAACAATA AGTTCCAGAC CCTCCTGGTA AGTCAGCCGA GGCAATAGCG TCATTTCGCG
 Ugd4   AAAAACAATA AGTTCCAGAC CCTCCTGGTA AGTCAGCCGA GGCAATAGCG TCATTTCGCG
Ugd16   AAAAACAATA AGTTCCAGAC CCTCCTGGTA AGTCAGCCGA GGCAATAGCG TCATTTCGCG
Ugd29   AAAAACAATA AGTTCCAGAC CCTCCTGGTA AGTCAGCCGA GGCAATAGCG TCATTTCGCG
Ugd12   AAAAACAATC AGTTCCAGAC CCTCCTGGTA AGTCAACCGA GGCAATAGTG TCATTTCGCG
Ugd19   AAAAACAATC AGTTCCAGAC CCTCCTGGTA AGTCAACCGA GGCAATAGTG TCATTTCGCG
 Ugd2   AAAAACAATC AGTTCCAGAC CCTCCTGGTA AGTCAACCGA GGCAATAGTG TCATTTCGCG
Ugd15   AAAAACAATC AGTTCCAGAC CCTCCTGGTA AGTCAGCCGA GGCAATAGCG TCATTTCGCG
Ugd23   AAAAACAATC AGTTCCAGAC CCTCCTGGTA AGTCAGCCGA GGCAATAGCG TCATTTCGCG
 BC-1   AAAAACAATC AGTTCCAGAC CCTCCTGGTA AGTCAGCCGA GGCAATAGCG TCATTTCGCG
 Cons   AAAAACAAT- AGTTCCAGAC CCTCCTGGTA AGTCA-CCGA GGCAATAG-G TCATTTCGCG

        361                                                              * 420
Ugd10   CAAGGGTCGC CAGACCACGC GCGTGTTGTA TACGACGCCA CATATCTGAC AGGCAGTGTT
 Ugd4   CAAGGGTCGC CAGACCACGC GCGTGTTGTA TACGACGCCA CATATCTGAC AGGCAGTGTT
Ugd16   CAAGGGTCGC CAGACCACGC GCGTGTTGTA TACGACGCCA CATATCTGAC AGGCAGTGTT
Ugd29   CAAGGGTCGC CAGACCACGC GCGTGTTGTA TACGACGCCA CATATCTGAC AGGCAGTGTT
Ugd12   CAAGGGTCGC CAGACCACGC GCGTGTTGTA TACGACGCCA CATATCTGAC AGGCCGTGTT
Ugd19   CAAGGGTCGC CAGACCACGC GCGTGTTGTA TACGACGCCA CATATCTGAC AGGCCGTGTT
 Ugd2   CAAGGGTCGC CAGACCACGC GCGTGTTGTA TACGACGCCA CATATCTGAC AGGCCGTGTT
Ugd15   CAAGGGTCGC CAGACCACGC GCGTGTTGTA TACGACGCCA CATATCTGAC AGGCCGTGTT
Ugd23   CAAGGGTCGC CAGACCACGC GCGTGTTGTA TACGACGCCA CATATCTGAC AGGCCGTGTT
 BC-1   CAAGGGTCGC CAGACCACGC GCGTGTTGTA TACGACGCCA CATATCTGAC AGGCCGTGTT
 Cons   CAAGGGTCGC CAGACCACGC GCGTGTTGTA TACGACGCCA CATATCTGAC AGGC-GTGTT

        421    *                                                         480
Ugd10   TCTAGAGATA GTGAGCCAGG TGCTTAAACA ACTTCTATGG ACGTTCTCGA GCTCTCCTGT
 Ugd4   TCTAGAGATA GTGAGCCAGG TGCTTAAACA ACTTCTATGG ACGTTCTCGA GCTCTCCTGT
Ugd16   TCTAGAGATA GTGAGCCAGG TGCTTAAACA ACTTCTATGG ACGTTCTCGA GCTCTCCTGT
Ugd29   TCTAGAGATA GTGAGCCAGG TGCTTAAACA ACTTCTATGG ACGTTCTCGA GCTCTCCTGT
Ugd12   TCTAGAGATA GTGAGCCAGG TGCTTAAACA ACTTCTATGG ACGTTCTCGA GCTCTCCTGT
Ugd19   TCTAGAGATA GTGAGCCAGG TGCTTAAACA ACTTCTATGG ACGTTCTCGA GCTCTCCTGT
 Ugd2   TCTAGAGATA GTGAGCCAGG TGCTTAAACA ACTTCTATGG ACGTTCTCGA GCTCTCCTGT
Ugd15   TCTAGAGATA GTGAGCCAGG TGCTTAAACA ACTTCTATGG ACGTTCTCGA GCTCTCCTGT
Ugd23   TCTAGATATA GTGAGCCAGG TGCTTAAACA ACTTCTATGG ACGTTCTCGA GCTCTCCTGT
 BC-1   TCTAGAGATA GTGAGCCAGG TGCTTAAACA ACTTCTATGG ACGTTCTCGA GCTCTCCTGT
 Cons   TCTAGA-ATA GTGAGCCAGG TGCTTAAACA ACTTCTATGG ACGTTCTCGA GCTCTCCTGT

        481          *                        519
Ugd10   GCATCCACAG GCTCGAAATC TCTCATTTCC GAGCTCCTC
 Ugd4   GCATCCACAG GCTCGAAATC TCTCATTTCC GAGCTCCTC
Ugd16   GCATCCACAG GCTCGAAATC TCTCATTTCC GAGCTCCTC
Ugd29   GCATCCACAG GCTCGAAATC TCTCATTTCC GAGCTCCTC
Ugd12   GCATCCACAG GCTCTAAATC TCTCATTTCC GAGCTCCTC
Ugd19   GCATCCACAG GCTCTAAATC TCTCATTTCC GAGCTCCTC
 Ugd2   GCATCCACAG GCTCTAAATC TCTCATTTCC GAGCTCCTC
Ugd15   GCATCCACAG GCTCTAAATC TCTCATTTCC GAGCTCCTC
Ugd23   GCATCCACAG GCTCTAAATC TCTCATTTCC GAGCTCCTC
 BC-1   GCATCCACAG GCTCTAAATC TCTCATTTCC GAGCTCCTC
 Cons   GCATCCACAG GCTC-AAATC TCTCATTTCC GAGCTCCTC
```

B - ORF26

```
        1                                                           60
Ugd12   GGGAGCGTAC TGCCGCTCGG AGATTGCCAC CGTTTACAAA ATATACAGGC ATTGGGCCTG
Ugd19   GGGAGCGTAC TGCCGCTCGG AGATTGCCAC CGTTTACAAA ATATACAGGC ATTGGGCCTG
 Ugd2   GGGAGCGTAC TGCCGCTCGG AGATTGCCAC CGTTTACAAA ATATACAGGC ATTGGGCCTG
Ugd23   GGGAGCGTAC TGCCGCTCGG AGATTGCCAC CGTTTACAAA ATATACAGGC ATTGGGCCTG
Ugd10   GGGAGCGTAC TGCCGCTCGG AGATTGCCAC CGTTTACAAA ATATACAGGC ATTGGGCCTG
Ugd15   GGGAGCGTAC TGCCGCTCGG AGATTGCCAC CGTTTACAAA ATATACAGGC ATTGGGCCTG
Ugd16   GGGAGCGTAC TGCCGCTCGG AGATTGCCAC CGTTTACAAA ATATACAGGC ATTGGGCCTG
Ugd29   GGGAGCGTAC TGCCGCTCGG AGATTGCCAC CGTTTACAAA ATATACAGGC ATTGGGCCTG
 Ugd4   GGGAGCGTAC TGCCGCTCGG AGATTGCCAC CGTTTACAAA ATATACAGGC ATTGGGCCTG
 BC-1   GGGAGCGTAC TGCCGCTCGG AGATTGCCAC CGTTTACAAA ATATACAGGC ATTGGGCCTG
 Cons   GGGAGCGTAC TGCCGCTCGG AGATTGCCAC CGTTTACAAA ATATACAGGC ATTGGGCCTG

        61                                                          120
Ugd12   GGGTGCGTAT GCTCACGTGA GACATCTCCG GACTACATCC AAATTATGCA GTATCTATCC
Ugd19   GGGTGCGTAT GCTCACGTGA GACATCTCCG GACTACATCC AAATTATGCA GTATCTATCC
 Ugd2   GGGTGCGTAT GCTCACGTGA GACATCTCCG GACTACATCC AAATTATGCA GTATCTATCC
Ugd23   GGGTGCGTAT GCTCACGTGA GACATCTCCG GACTACATCC AAATTATGCA GTATCTATCC
Ugd10   GGGTGCGTAT GCTCACGTGA GACATCTCCG GACTACATCC AAATTATGCA GTATCTATCC
Ugd15   GGGTGCGTAT GCTCACGTGA GACATCTCCG GACTACATCC AAATTATGCA GTATCTATCC
Ugd16   GGGTGCGTAT GCTCACGTGA GACATCTCCG GACTACATCC AAATTATGCA GTATCTATCC
Ugd29   GGGTGCGTAT GCTCACGTGA GACATCTCCG GACTACATCC AAATTATGCA GTATCTATCC
 Ugd4   GGGTGCGTAT GCTCACGTGA GACATCTCCG GACTACATCC AAATTATGCA GTATCTATCC
 BC-1   GGGTGCGTAT GCTCACGTGA GACATCTCCG GACTACATCC AAATTATGCA GTATCTATCC
 Cons   GGGTGCGTAT GCTCACGTGA GACATCTCCG GACTACATCC AAATTATGCA GTATCTATCC

        121                                                         180
Ugd12   AAGTGCACAC TCGCTGTCCT GGAGGAGGTT CGCCCGGACA GCCTGCGCCT AACGCGGATG
Ugd19   AAGTGCACAC TCGCTGTCCT GGAGGAGGTT CGCCCGGACA GCCTGCGCCT AACGCGGATG
 Ugd2   AAGTGCACAC TCGCTGTCCT GGAGGAGGTT CGCCCGGACA GCCTGCGCCT AACGCGGATG
Ugd23   AAGTGCACAC TCGCTGTCCT GGAGGAGGTT CGCCCGGACA GCCTGCGCCT AACGCGGATG
Ugd10   AAGTGCACAC TCGCTGTCCT GGAGGAGGTT CGCCCGGACA GCCTGCGCCT AACGCGGATG
Ugd15   AAGTGCACAC TCGCTGTCCT GGAGGAGGTT CGCCCGGACA GCCTGCGCCT AACGCGGATG
Ugd16   AAGTGCACAC TCGCTGTCCT GGAGGAGGTT CGCCCGGACA GCCTGCGCCT AACGCGGATG
Ugd29   AAGTGCACAC TCGCTGTCCT GGAGGAGGTT CGCCCGGACA GCCTGCGCCT AACGCGGATG
 Ugd4   AAGTGCACAC TCGCTGTCCT GGAGGAGGTT CGCCCGGACA GCCTGCGCCT AACGCGGATG
 BC-1   AAGTGCACAC TCGCTGTCCT GGAGGAGGTT CGCCCGGACA GCCTGCGCCT AACGCGGATG
 Cons   AAGTGCACAC TCGCTGTCCT GGAGGAGGTT CGCCCGGACA GCCTGCGCCT AACGCGGATG

        181                                                         240
Ugd12   GATCCCTCTG ACAACCTTCA GATAAAAAAC GTATATGCCC CCTTTTTTCA GTGGGACAGC
Ugd19   GATCCCTCTG ACAACCTTCA GATAAAAAAC GTATATGCCC CCTTTTTTCA GTGGGACAGC
 Ugd2   GATCCCTCTG ACAACCTTCA GATAAAAAAC GTATATGCCC CCTTTTTTCA GTGGGACAGC
Ugd23   GATCCCTCTG ACAACCTTCA GATAAAAAAC GTATATGCCC CCTTTTTTCA GTGGGACAGC
Ugd10   GATCCCTCTG ACAACCTTCA GATAAAAAAC GTATATGCCC CCTTTTTTCA GTGGGACAGC
Ugd15   GATCCCTCTG ACAACCTTCA GATAAAAAAC GTATATGCCC CCTTTTTTCA GTGGGACAGC
Ugd16   GATCCCTCTG ACAACCTTCA GATAAAAAAC GTATATGCCC CCTTTTTTCA GTGGGACAGC
Ugd29   GATCCCTCTG ACAACCTTCA GATAAAAAAC GTATATGCCC CCTTTTTTCA GTGGGACAGC
 Ugd4   GATCCCTCTG ACAACCTTCA GATAAAAAAC GTATATGCCC CCTTTTTTCA GTGGGACAGC
 BC-1   GATCCCTCTG ACAACCTTCA GATAAAAAAC GTATATGCCC CCTTTTTTCA GTGGGACAGC
 Cons   GATCCCTCTG ACAACCTTCA GATAAAAAAC GTATATGCCC CCTTTTTTCA GTGGGACAGC

                                    L/F
        241                                                         300
Ugd12   AACACCCAGC TAGCAGTGCT ACCCCCACTT TTTAGCCGAA AGGATTCCAC CATTGTGCTC
Ugd19   AACACCCAGC TAGCAGTGCT ACCCCCACTT TTTAGCCGAA AGGATTCCAC CATTGTGCTC
 Ugd2   AACACCCAGC TAGCAGTGCT ACCCCCACTT TTTAGCCGAA AGGATTCCAC CATTGTGCTC
Ugd23   AACACCCAGC TAGCAGTGCT ACCCCCACTT TTTAGCCGAA AGGATTCCAC CATTGTGCTC
Ugd10   AACACCCAGC TAGCAGTGCT ACCCCCACTT TTTAGCCGAA AGGATTCCAC CATTGTGCTC
Ugd15   AACACCCAGC TAGCAGTGCT ACCCCCACTT TTTAGCCGAA AGGATTCCAC CATTGTGCTC
Ugd16   AACACCCAGC TAGCAGTGCT ACCCCCACTT TTTAGCCGAA AGGATTCCAC CATTGTGCTC
Ugd29   AACACCCAGC TAGCAGTGCT ACCCCCACTT TTTAGCCGAA AGGATTCCAC CATTGTGCTC
 Ugd4   AACACCCAGC TAGCAGTGCT ACCCCCACTT TTTAGCCGAA AGGATTCCAC CATTGTGCTC
 BC-1   AACACCCAGC TAGCAGTGCT ACCCCCATTT TTTAGCCGAA AGGATTCCAC CATTGTGCTC
 Cons   AACACCCAGC TAGCAGTGCT ACCCCCA-TT TTTAGCCGAA AGGATTCCAC CATTGTGCTC
```

```
        301                  I/L                                *                  360
Ugd12   GAATCCAACG GATTTGACAT CGTGTTCCCC ATGGTCGTGC CGCAGCAACT GGGGCACGCT
Ugd19   GAATCCAACG GATTTGACAT CGTGTTCCCC ATGGTCGTGC CGCAGCAACT GGGGCACGCT
 Ugd2   GAATCCAACG GATTTGACAT CGTGTTCCCC ATGGTCGTGC CGCAGCAACT GGGGCACGCT
Ugd23   GAATCCAACG GATTTGACAT CGTGTTCCCC ATGGTCGTGC CTCAGCAACT GGGGCACGCT
Ugd10   GAATCCAACG GATTTGACCT CGTGTTCCCC ATGGTCGTGC CGCAGCAACT GGGGCACGCT
Ugd15   GAATCCAACG GATTTGACCT CGTGTTCCCC ATGGTCGTGC CGCAGCAACT GGGGCACGCT
Ugd16   GAATCCAACG GATTTGACCT CGTGTTCCCC ATGGTCGTGC CGCAGCAACT GGGGCACGCT
Ugd29   GAATCCAACG GATTTGACCT CGTGTTCCCC ATGGTCGTGC CGCAGCAACT GGGGCACGCT
 Ugd4   GAATCCAACG GATTTGACCT CGTGTTCCCC ATGGTCGTGC CGCAGCAACT GGGGCACGCT
 BC-1   GAATCCAACG GATTTGACCT CGTGTTCCCC ATGGTCGTGC CGCAGCAACT GGGGCACGCT
 Cons   GAATCCAACG GATTTGAC-T CGTGTTCCCC ATGGTCGTGC C-CAGCAACT GGGGCACGCT
                                                                       G/M
        361             *                                                  420
Ugd12   ATTCTGCAGC AGCTGTTGGT GTACCACATC TACTCCAAAA TATCGGCCGG GGCCCCGGGT
Ugd19   ATTCTGCAGC AGCTGTTGGT GTACCACATC TACTCCAAAA TATCGGCCGG GGCCCCGGGT
 Ugd2   ATTCTGCAGC AGCTGTTGGT GTACCACATC TACTCCAAAA TATCGGCCGG GGCCCCGGGT
Ugd23   ATTCTGCAGC AGCTGTTGGT GTACCACATC TACTCCAAAA TATCGGCCGG GGCCCCGGGT
Ugd10   ATTCTGCAGC AGTTGTTGGT GTACCACATC TACTCCAAAA TATCGGCCGG GGCCCCGGAT
Ugd15   ATTCTGCAGC AGTTGTTGGT GTACCACATC TACTCCAAAA TATCGGCCGG GGCCCCGGAT
Ugd16   ATTCTGCAGC AGTTGTTGGT GTACCACATC TACTCCAAAA TATCGGCCGG GGCCCCGGAT
Ugd29   ATTCTGCAGC AGTTGTTGGT GTACCACATC TACTCCAAAA TATCGGCCGG GGCCCCGGAT
 Ugd4   ATTCTGCAGC AGTTGTTGGT GTACCACATC TACTCCAAAA TATCGGCCGG GGCCCCGGAT
 BC-1   ATTCTGCAGC AGCTGTTGGT GTACCACATC TACTCCAAAA TATCGGCCGG GGCCCCGGAT
 Cons   ATTCTGCAGC AG-TGTTGGT GTACCACATC TACTCCAAAA TATCGGCCGG GGCCCCGG-T

        421 *                                                              480
Ugd12   GATGTCAATA TGGCGGAACT TGATCTATAT ACCACCAATG TGTCATTTAT GGGGCGCACA
Ugd19   GATGTCAATA TGGCGGAACT TGATCTATAT ACCACCAATG TGTCATTTAT GGGGCGCACA
 Ugd2   GATGTCAATA TGGCGGAACT TGATCTATAT ACCACCAATG TGTCATTTAT GGGGCGCACA
Ugd23   GATGTCAATA TGGCGGAACT TGATCTATAT ACCACCAATG TGTCATTTAT GGGGCGCACA
Ugd10   GATGTCAATA TGGCGGAACT TGATCTATAT ACCACCAATG TGTCATTTAT GGGGCGCACA
Ugd15   GATGTCAATA TGGCGGAACT TGATCTATAT ACCACCAATG TGTCATTTAT GGGGCGCACA
Ugd16   GATGTCAATA TGGCGGAACT TGATCTATAT ACCACCAATG TGTCATTTAT GGGGCGCACA
Ugd29   GATGTCAATA TGGCGGAACT TGATCTATAT ACCACCAATG TGTCATTTAT GGGGCGCACA
 Ugd4   GATGTCAATA TGGCGGAACT TGATCTATAT ACCACCAATG TGTCATTTAT GGGGCGCACA
 BC-1   GATGTAAATA TGGCGGAACT TGATCTATAT ACCACCAATG TGTCATTTAT GGGGCGCACA
 Cons   GATGT-AATA TGGCGGAACT TGATCTATAT ACCACCAATG TGTCATTTAT GGGGCGCACA

        481          495
Ugd12   TATCGTCTGG ACGTA
Ugd19   TATCGTCTGG ACGTA
 Ugd2   TATCGTCTGG ACGTA
Ugd23   TATCGTCTGG ACGTA
Ugd10   TATCGTCTGG ACGTA
Ugd15   TATCGTCTGG ACGTA
Ugd16   TATCGTCTGG ACGTA
Ugd29   TATCGTCTGG ACGTA
 Ugd4   TATCGTCTGG ACGTA
 BC-1   TATCGTCTGG ACGTA
 Cons   TATCGTCTGG ACGTA
```

C - K9

```
          1                                                                 60
 BC-1     ATTGACTGGG GTCGGTTGTT TATCAGGATG TATTATAATG GCGAACAGGT TCATGAGTTA
Ugd10     ATTGACTGGG GTCGGTTGTT TATCAGGATG TATTATAATG GCGAACAGGT TCATGAGTTA
Ugd12     ATTGACTGGG GTCGGTTGTT TATCAGGATG TATTATAATG GCGAACAGGT TCATGAGTTA
Ugd15     ATTGACTGGG GTCGGTTGTT TATCAGGATG TATTATAATG GCGAACAGGT TCATGAGTTA
Ugd19     ATTGACTGGG GTCGGTTGTT TATCAGGATG TATTATAATG GCGAACAGGT TCATGAGTTA
 Ugd2     ATTGACTGGG GTCGGTTGTT TATCAGGATG TATTATAATG GCGAACAGGT TCATGAGTTA
Ugd23     ATTGACTGGG GTCGGTTGTT TATCAGGATG TATTATAATG GCGAACAGGT TCATGAGTTA
Ugd29     ATTGACTGGG GTCGGTTGTT TATCAGGATG TATTATAATG GCGAACAGGT TCATGAGTTA
Ugd16     ATTGACTGGG GTCGGTTGTT TATCAGGATG TATTATAATG GCGAACAGGT TCATGAGTTA
 Ugd4     ATTGACTGGG GTCGGTTGTT TATCAGGATG TATTATAATG GCGAACAGGT TCATGAGTTA
 Cons     ATTGACTGGG GTCGGTTGTT TATCAGGATG TATTATAATG GCGAACAGGT TCATGAGTTA

          61                                                               120
 BC-1     CTGACGACAA GCCAGTCGGG CTGTCGAATC TCATCGGCCT TGCGTAGGGA CCCCGCAGTT
Ugd10     CTGACGACAA GCCAGTCGGG CTGTCGAATC TCATCGGCCT TGCGTAGGGA CCCCGCAGTT
Ugd12     CTGACGACAA GCCAGTCGGG CTGTCGAATC TCATCGGCCT TGCGTAGGGA CCCCGCAGTT
Ugd15     CTGACGACAA GCCAGTCGGG CTGTCGAATC TCATCGGCCT TGCGTAGGGA CCCCGCAGTT
Ugd19     CTGACGACAA GCCAGTCGGG CTGTCGAATC TCATCGGCCT TGCGTAGGGA CCCCGCAGTT
 Ugd2     CTGACGACAA GCCAGTCGGG CTGTCGAATC TCATCGGCCT TGCGTAGGGA CCCCGCAGTT
Ugd23     CTGACGACAA GCCAGTCGGG CTGTCGAATC TCATCGGCCT TGCGTAGGGA CCCCGCAGTT
Ugd29     CTGACGACAA GCCAGTCGGG CTGTCGAATC TCATCGGCCT TGCGTAGGGA CCCCGCAGTT
Ugd16     CTGACGACAA GCCAGTCGGG CTGTCGAATC TCATCGGCCT TGCGTAGGGA CCCCGCAGTT
 Ugd4     CTGACGACAA GCCAGTCGGG CTGTCGAATC TCATCGGCCT TGCGTAGGGA CCCCGCAGTT
 Cons     CTGACGACAA GCCAGTCGGG CTGTCGAATC TCATCGGCCT TGCGTAGGGA CCCCGCAGTT

          121                                                              180
 BC-1     CATTACTGTG CAGTGGGGTC TCCGGGCCAG GTATGGCTAC CCAATGTGCC AAACCTGGCC
Ugd10     CATTACTGTG CAGTGGGGTC TCCGGGCCAG GTATGGCTAC CCAATGTGCC AAACCTGGCC
Ugd12     CATTACTGTG CAGTGGGGTC TCCGGGCCAG GTATGGCTAC CCAATGTGCC AAACCTGGCC
Ugd15     CATTACTGTG CAGTGGGGTC TCCGGGCCAG GTATGGCTAC CCAATGTGCC AAACCTGGCC
Ugd19     CATTACTGTG CAGTGGGGTC TCCGGGCCAG GTATGGCTAC CCAATGTGCC AAACCTGGCC
 Ugd2     CATTACTGTG CAGTGGGGTC TCCGGGCCAG GTATGGCTAC CCAATGTGCC AAACCTGGCC
Ugd23     CATTACTGTG CAGTGGGGTC TCCGGGCCAG GTATGGCTAC CCAATGTGCC AAACCTGGCC
Ugd29     CATTACTGTG CAGTGGGGTC TCCGGGCCAG GTATGGCTAC CCAATGTGCC AAACCTGGCC
Ugd16     CATTACTGTG CAGTGGGGTC TCCGGGCCAG GTATGGCTAC CCAATGTGCC AAACCTGGCC
 Ugd4     CATTACTGTG CAGTGGGGTC TCCGGGCCAG GTATGGCTAC CCAATGTGCC AAACCTGGCC
 Cons     CATTACTGTG CAGTGGGGTC TCCGGGCCAG GTATGGCTAC CCAATGTGCC AAACCTGGCC

          181                                                              240
 BC-1     TGCGAGATAG CCAAGCGGGA GCTATGCGAT ACGCTGGACG CATGTGCAAA AGGCATTCTG
Ugd10     TGCGAGATAG CCAAGCGGGA GCTATGCGAT ACGCTGGACG CATGTGCAAA AGGCATTCTG
Ugd12     TGCGAGATAG CCAAGCGGGA GCTATGCGAT ACGCTGGACG CATGTGCAAA AGGCATTCTG
Ugd15     TGCGAGATAG CCAAGCGGGA GCTATGCGAT ACGCTGGACG CATGTGCAAA AGGCATTCTG
Ugd19     TGCGAGATAG CCAAGCGGGA GCTATGCGAT ACGCTGGACG CATGTGCAAA AGGCATTCTG
 Ugd2     TGCGAGATAG CCAAGCGGGA GCTATGCGAT ACGCTGGACG CATGTGCAAA AGGCATTCTG
Ugd23     TGCGAGATAG CCAAGCGGGA GCTATGCGAT ACGCTGGACG CATGTGCAAA AGGCATTCTG
Ugd29     TGCGAGATAG CCAAGCGGGA GCTATGCGAT ACGCTGGACG CATGTGCAAA AGGCATTCTG
Ugd16     TGCGAGATAG CCAAGCGGGA GCTATGCGAT ACGCTGGACG CATGTGCAAA AGGCATTCTG
 Ugd4     TGCGAGATAG CCAAGCGGGA GCTATGCGAT ACGCTGGACG CATGTGCAAA AGGCATTCTG
 Cons     TGCGAGATAG CCAAGCGGGA GCTATGCGAT ACGCTGGACG CATGTGCAAA AGGCATTCTG

          241                                                              300
 BC-1     CTGACTAGCT CTTGTAATGG CATATTTTGC GTATGTTATC ATAACGGGCC CGTGCACTTT
Ugd10     CTGACTAGCT CTTGTAATGG CATATTTTGC GTATGTTATC ATAACGGGCC CGTGCACTTT
Ugd12     CTGACTAGCT CTTGTAATGG CATATTTTGC GTATGTTATC ATAACGGGCC CGTGCACTTT
Ugd15     CTGACTAGCT CTTGTAATGG CATATTTTGC GTATGTTATC ATAACGGGCC CGTGCACTTT
Ugd19     CTGACTAGCT CTTGTAATGG CATATTTTGC GTATGTTATC ATAACGGGCC CGTGCACTTT
 Ugd2     CTGACTAGCT CTTGTAATGG CATATTTTGC GTATGTTATC ATAACGGGCC CGTGCACTTT
Ugd23     CTGACTAGCT CTTGTAATGG CATATTTTGC GTATGTTATC ATAACGGGCC CGTGCACTTT
Ugd29     CTGACTAGCT CTTGTAATGG CATATTTTGC GTATGTTATC ATAACGGGCC CGTGCACTTT
Ugd16     CTGACTAGCT CTTGTAATGG CATATTTTGC GTATGTTATC ATAACGGGCC CGTGCACTTT
 Ugd4     CTGACTAGCT CTTGTAATGG CATATTTTGC GTATGTTATC ATAACGGGCC CGTGCACTTT
 Cons     CTGACTAGCT CTTGTAATGG CATATTTTGC GTATGTTATC ATAACGGGCC CGTGCACTTT
```

```
        301                                                     *     360
 BC-1   ATTGGAAATA CCGTTCCGCC TGACTCCGGC CCCCTCTTGC TGCCCCAGGG AAAACCGACG
Ugd10   ATTGGAAATA CCGTTCCGCC TGACTCCGGC CCCCTCTTGC TGCCCCAGGG AAAACCGACG
Ugd12   ATTGGAAATA CCGTTCCGCC TGACTCCGGC CCCCTCTTGC TGCCCCAGGG AAAACCGACG
Ugd15   ATTGGAAATA CCGTTCCGCC TGACTCCGGC CCCCTCTTGC TGCCCCAGGG AAAACCGACG
Ugd19   ATTGGAAATA CCGTTCCGCC TGACTCCGGC CCCCTCTTGC TGCCCCAGGG AAAACCGACG
 Ugd2   ATTGGAAATA CCGTTCCGCC TGACTCCGGC CCCCTCTTGC TGCCCCAGGG AAAACCGACG
Ugd23   ATTGGAAATA CCGTTCCGCC TGACTCCGGC CCCCTCTTGC TGCCCCAGGG AAAACCGACG
Ugd29   ATTGGAAATA CCGTTCCGCC TGACTCCGGC CCCCTCTTGC TGCCCCAGGG AAAACCGACG
Ugd16   ATTGGAAATA CCGTTCCGCC TGACTCCGGC CCCCTCTTGC TGCCCCAGGG AAAGCCGACG
 Ugd4   ATTGGAAATA CCGTTCCGCC TGACTCCGGC CCCCTCTTGC TGCCCCAGGG AAAGCCGACG
 Cons   ATTGGAAATA CCGTTCCGCC TGACTCCGGC CCCCTCTTGC TGCCCCAGGG AAA-CCGACG

        361                                                         420
 BC-1   AGGATATTTA ACCCGAATAC ATTTCTGGTG GGGTTGGCAA ATTCACCGTT ACCAGCACCC
Ugd10   AGGATATTTA ACCCGAATAC ATTTCTGGTG GGGTTGGCAA ATTCACCGTT ACCAGCACCC
Ugd12   AGGATATTTA ACCCGAATAC ATTTCTGGTG GGGTTGGCAA ATTCACCGTT ACCAGCACCC
Ugd15   AGGATATTTA ACCCGAATAC ATTTCTGGTG GGGTTGGCAA ATTCACCGTT ACCAGCACCC
Ugd19   AGGATATTTA ACCCGAATAC ATTTCTGGTG GGGTTGGCAA ATTCACCGTT ACCAGCACCC
 Ugd2   AGGATATTTA ACCCGAATAC ATTTCTGGTG GGGTTGGCAA ATTCACCGTT ACCAGCACCC
Ugd23   AGGATATTTA ACCCGAATAC ATTTCTGGTG GGGTTGGCAA ATTCACCGTT ACCAGCACCC
Ugd29   AGGATATTTA ACCCGAATAC ATTTCTGGTG GGGTTGGCAA ATTCACCGTT ACCAGCACCC
Ugd16   AGGATATTTA ACCCGAATAC ATTTCTGGTG GGGTTGGCAA ATTCACCGTT ACCAGCACCC
 Ugd4   AGGATATTTA ACCCGAATAC ATTTCTGGTG GGGTTGGCAA ATTCACCGTT ACCAGCACCC
 Cons   AGGATATTTA ACCCGAATAC ATTTCTGGTG GGGTTGGCAA ATTCACCGTT ACCAGCACCC

        421                                                         480
 BC-1   TCTCACGTGA CGTGTCCTTT GGTGAAGCTG TGGCTGGGGA AACCGGTGGC GGTTGGCAAG
Ugd10   TCTCACGTGA CGTGTCCTTT GGTGAAGCTG TGGCTGGGGA AACCGGTGGC GGTTGGCAAG
Ugd12   TCTCACGTGA CGTGTCCTTT GGTGAAGCTG TGGCTGGGGA AACCGGTGGC GGTTGGCAAG
Ugd15   TCTCACGTGA CGTGTCCTTT GGTGAAGCTG TGGCTGGGGA AACCGGTGGC GGTTGGCAAG
Ugd19   TCTCACGTGA CGTGTCCTTT GGTGAAGCTG TGGCTGGGGA AACCGGTGGC GGTTGGCAAG
 Ugd2   TCTCACGTGA CGTGTCCTTT GGTGAAGCTG TGGCTGGGGA AACCGGTGGC GGTTGGCAAG
Ugd23   TCTCACGTGA CGTGTCCTTT GGTGAAGCTG TGGCTGGGGA AACCGGTGGC GGTTGGCAAG
Ugd29   TCTCACGTGA CGTGTCCTTT GGTGAAGCTG TGGCTGGGGA AACCGGTGGC GGTTGGCAAG
Ugd16   TCTCACGTGA CGTGTCCTTT GGTGAAGCTG TGGCTGGGGA AACCGGTGGC GGTTGGCAAG
 Ugd4   TCTCACGTGA CGTGTCCTTT GGTGAAGCTG TGGCTGGGGA AACCGGTGGC GGTTGGCAAG
 Cons   TCTCACGTGA CGTGTCCTTT GGTGAAGCTG TGGCTGGGGA AACCGGTGGC GGTTGGCAAG

        481             500
 BC-1   TTAGAGCCCC ACGCCCCGTC
Ugd10   TTAGAGCCCC ACGCCCCGTC
Ugd12   TTAGAGCCCC ACGCCCCGTC
Ugd15   TTAGAGCCCC ACGCCCCGTC
Ugd19   TTAGAGCCCC ACGCCCCGTC
 Ugd2   TTAGAGCCCC ACGCCCCGTC
Ugd23   TTAGAGCCCC ACGCCCCGTC
Ugd29   TTAGAGCCCC ACGCCCCGTC
Ugd16   TTAGAGCCCC ACGCCCCGTC
 Ugd4   TTAGAGCCCC ACGCCCCGTC
 Cons   TTAGAGCCCC ACGCCCCGTC
```

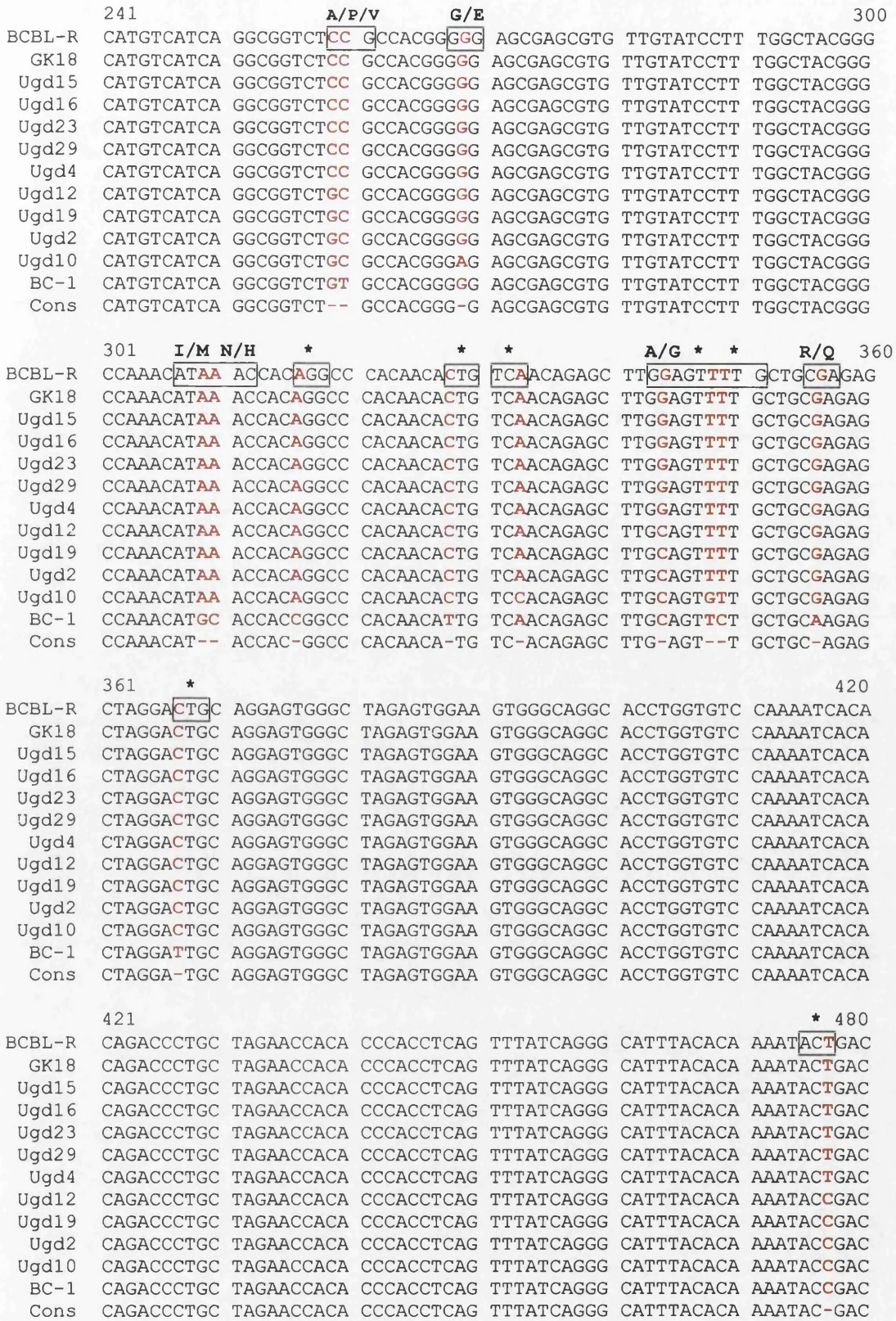
D - T0.7/K12

```
        1  P/L                                          *                    60
Ugd19   GTCCCGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
 Ugd2   GTCCCGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTCG GGGCGATACC ACCACTCGTT
Ugd12   GTCCCGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
Ugd16   GTCCCGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
Ugd29   GTCCCGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
Ugd23   GTCCCGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
 Ugd4   GTCCTGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
 BC-1   GTCCCGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
Ugd10   GTCCCGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
Ugd15   GTCCCGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
 Cons   GTCC-GGATG TGTTACTAAA TGGGTGGCGC TGGAGGCT-G GGGCGATACC ACCACTCGTT

        61 L/V                          *                          V/E A/P *  120
Ugd19   TGTCTGTTGG CGATTAGTGT TGTCCCCCCG AGTGGCCAGC GTGGCCCCGT ACCATTCAGG
 Ugd2   TGTCTGTTGG CGATTAGTGT TGTCCCCCCG AGTGGCCAGC GTGGCCCCGT ACCATTCAGG
Ugd12   TGTCTGTTGG CGATTAGTGT TGTCCCCCCA AGTGGCCAGC GTGGCCCCGT ACCATTTAGG
Ugd16   TGTGTGTTGG CGATTAGTGT TGTCCCCCCA AGTGGCCAGC GTGGCCCCGT ACCATTCAGG
Ugd29   TGTGTGTTGG CGATTAGTGT TGTCCCCCCA AGTGGCCAGC GTGGCCCCGT ACCATTCAGG
Ugd23   TGTCTGTTGG CGATTAGTGT TGTCCCCCCG AGTGGCCAGC GTGGCCCCGT ACCATTCAGG
 Ugd4   TGTGTGTTGG CGATTAGTGT TGTCCCCCCG AGTGGCCAGC GTGGCCCCGT ACCATTCAGG
 BC-1   TGTCTGTTGG CGATTAGTGT TGTCCCCCCG AGTGGCCAGC GTGGCCCCGT AGCATTCAGG
Ugd10   TGTCTGTTGG CGATTAGTGT TGTCCCCCCA AGTGGCCAGC GTGGCCCCGT AGCATTCAGG
Ugd15   TGTCTGTTGG CGATTAGTGT TGTCCCCCCG AGTGGCCAGC GTGGCCCCGA ACCATTCAGG
 Cons   TGT-TGTTGG CGATTAGTGT TGTCCCCCC- AGTGGCCAGC GTGGCCCCG- A-CATT-AGG

        121                          stop                                180
Ugd19   ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG... CCCTCCCGAA GAGTGTCAGT
 Ugd2   ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG... CCCTCCCGAA GAGTGTCAGT
Ugd12   ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG... CCCTCCCGAA GAGTGTCAGT
Ugd16   ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG... CCCTCCCGAG GAGTGTCAGT
Ugd29   ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG... CCCTCTCGAG GAGTGTCAGT
Ugd23   ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG... CCCTCCCGAG GAGTGTCAGT
 Ugd4   ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG... CCCTCCCGAG GAGTGTCAGT
 BC-1   ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG... CCCTCCCAAA GAGTGTCAGT
Ugd10   ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG... CCCTCCCGAA GAGTGTCAGT
Ugd15   ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTGCCC CCCCCCCGAG GAGTGTCAGT
 Cons   ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG--- CCC-C-C-A- GAGTGTCAGT

        181                                                          240
Ugd19   AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATCAAA CACATACAAT
 Ugd2   AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATCAAA CACATACAAT
Ugd12   AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATCAAA CACATACAAT
Ugd16   AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATCAAA CACATACAAT
Ugd29   AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATCAAA CACATACAAT
Ugd23   AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATCAAA CACATACAAT
 Ugd4   AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATCAAA CACATACAAT
 BC-1   AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATTAAA CACATACAAT
Ugd10   AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATCAAA CACATACAAT
Ugd15   AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATCAAA CACATACAAT
 Cons   AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACAT-AAA CACATACAAT

        241                                                          300
Ugd19   GCTGAAGAGT AGGAGTATCG AGGGCATATC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
 Ugd2   GCTGAAGAGT AGGAGTATCG AGGGCATATC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
Ugd12   GCTGAAGAGT AGGAGTATCG AGGGCATATC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
Ugd16   GCTGAAGAGC AGGCGTATCG AGGGCATACC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
Ugd29   GCTGAAGAGC AGGCGTATCG AGGGCATACC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
Ugd23   GCTGAAGAGC AGGCGTATCG AGGGCATACC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
 Ugd4   GCTGAAGAGC AGGCGTATCG AGGGCATACC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
 BC-1   GCTGAAGAGC AGGCGTATCG AGGGCATACC CGTGGCTCAT AACACAGTCA CAGTTCAGAT
Ugd10   CCTGAAGAGC AGGCGTATCG AGGGCATACC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
Ugd15   GCTGAAGAGC AGGCGTATCG AGGGCATACC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
 Cons   -CTGAAGAG- AGG-GTATCG AGGGCATA-C CGTGGCTCAT AACACAGTCA CAGTTCAGA-
```

```
        301                                                         360
Ugd19   GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
 Ugd2   GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
Ugd12   GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
Ugd16   GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
Ugd29   GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
Ugd23   GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
 Ugd4   GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
 BC-1   GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
Ugd10   GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
Ugd15   GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
 Cons   GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG

        361                                                         420
Ugd19   GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
 Ugd2   GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
Ugd12   GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
Ugd16   GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
Ugd29   GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
Ugd23   GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
 Ugd4   GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
 BC-1   GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
Ugd10   GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
Ugd15   GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
 Cons   GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC

        421                                                         480
Ugd19   CAGCAAAGCA TATTTTGCAC GCAAATGGTG TCCGCTCTCC CAAACCACAC GAATGGCACC
 Ugd2   CAGCAAAGCA TATTTTGCAC GCAAATGGTG TCCGCTCTCC CAAACCACAC GAATGGCACC
Ugd12   CAGCAAAGCA TATTTTGCAC GCAAATGGTG TCCGCTCTCC CAAACCACAC GAATGGCACC
Ugd16   CAGCAAAGCA CATTTTGCAC GCAAATGGTG TCCGCTCTCC CAAACCACAC GAATGGCACC
Ugd29   CAGCAAAGCA CATTTTGCAC GCAAATGGTG TCCGCTCTCC CAAACCACAC GAATGGCACC
Ugd23   CAGCAAAGCA CATTTTGCAC GCAAATGGTG TCCGCTCTCC CAAACCACAC GAATGGCACC
 Ugd4   CAGCAAAGCA CATTTTGCAC GCAAATGGTG TCCGCTCTCC CAAACCACAC GAATGGCACC
 BC-1   CAGCAAAGCA CATTTTGCAC GCAAATGGCG TCCGCTCTCC CAAACCACAC GAATGGTACC
Ugd10   CAGCAAAGCA CATTTTGCAC GCAAATGGCG TCCGCTCTCC CAAACCACAC GGATGGCACC
Ugd15   CAGCAAAGCA CATTTTGCAC GCAAATGGTG TCCGCTCTCC CAAACCACAC GAATGGCACC
 Cons   CAGCAAAGCA -ATTTTGCAC GCAAATGG-G TCCGCTCTCC CAAACCACAC G-ATGG-ACC

        481                                                         540
Ugd19   ATGGCAAAAA ACACTCCCCC ACAATGTATT GACCTTGTTT GTTAATGAAA TTGTGGCACT
 Ugd2   ATGGCAAAAA ACACTCCCCC ACAATGTATT GACCTTGTTT GTTAATGAAA TTGTGGCACT
Ugd12   ATGGCAAAAA ACACTCCCCC ACAATGTATT GACCTTGTTT GTTAATGAAA TTGTGGCACT
Ugd16   ATGGCAAAAA ACACTCCCCC ACAATGTATT GACCTTGTTT GTTAATGAAA GTGTGGCATT
Ugd29   ATGGCAAAAA ACACTCCCCC ACAATGTATT GACCTTGTTT GTTAATGAAA GTGTGGCACT
Ugd23   ATGGCAAAAA ACACTCCCCC ACAATGTATT GACCTTGTTT GTTAATGAAA GTGTGGCACT
 Ugd4   ATGGCAAAAA ACACTCCCCC ACAATGTATT GACCTTGTTT GTTAATGAAA GTGTGGCATT
 BC-1   ATGGCAAAAA ACACTCCCTC ACAATGTATT GACCTTGTTT GTTAATGAAA TTGTGGCACT
Ugd10   ATGGCAAAAA ACCCTCCCCC ACAATGTATT GACCTTGTTT GTTAATGAAA TTGTGGCACT
Ugd15   ATGGCAAAAA ACACTCCCCC ACAATGTATT GACCTTGTTT GTTAATGAAA GTGTGGCACT
 Cons   ATGGCAAAAA AC-CTCCC-C ACAATGTATT GACCTTGTTT GTTAATGAAA -TGTGGCA-T

        541      551
Ugd19   GACTTCGGCA G
 Ugd2   GACTTCGGCA G
Ugd12   GACTTCGGCA G
Ugd16   GACTTCGGCA G
Ugd29   GACTTCGGCA G
Ugd23   GACTTCGGCA G
 Ugd4   GACTTCGGCA G
 BC-1   GACTTCGGCA G
Ugd10   GACTTCGGCA G
Ugd15   GACTTCGGCA G
 Cons   GACTTCGGCA G
```

E - ORF75

```
          1start                                                                 60
BCBL-R    ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
  GK18    ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
 Ugd15    ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
 Ugd16    ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
 Ugd23    ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
 Ugd29    ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
  Ugd4    ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
 Ugd12    ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
 Ugd19    ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
  Ugd2    ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
 Ugd10    ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
  BC-1    ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
  Cons    ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT

          61  * S/N                                                          120
BCBL-R    TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
  GK18    TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
 Ugd15    TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
 Ugd16    TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
 Ugd23    TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
 Ugd29    TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
  Ugd4    TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
 Ugd12    TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
 Ugd19    TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
  Ugd2    TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
 Ugd10    TTTGTGAACT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
  BC-1    TTTGTGAACT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
  Cons    TTTGT-A-CT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA

          121                                                            180
BCBL-R    GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
  GK18    GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
 Ugd15    GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
 Ugd16    GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
 Ugd23    GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
 Ugd29    GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
  Ugd4    GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
 Ugd12    GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
 Ugd19    GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
  Ugd2    GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
 Ugd10    GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
  BC-1    GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
  Cons    GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG

          181                                                            240
BCBL-R    AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
  GK18    AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
 Ugd15    AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
 Ugd16    AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
 Ugd23    AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
 Ugd29    AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
  Ugd4    AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
 Ugd12    AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
 Ugd19    AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
  Ugd2    AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
 Ugd10    AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
  BC-1    AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
  Cons    AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
```

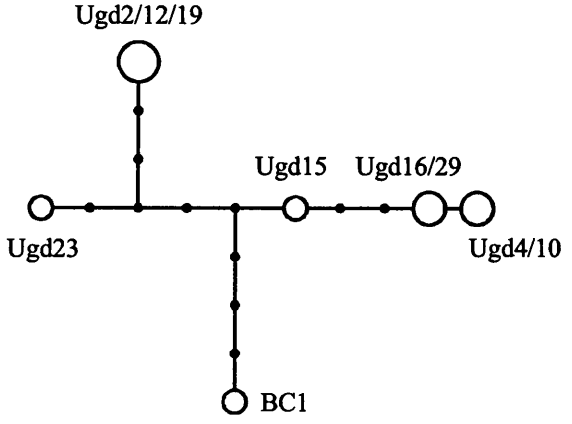

241 A/P/V G/E 300
BCBL-R CATGTCATCA GGCGGTCTCC GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
GK18 CATGTCATCA GGCGGTCTCC GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
Ugd15 CATGTCATCA GGCGGTCTCC GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
Ugd16 CATGTCATCA GGCGGTCTCC GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
Ugd23 CATGTCATCA GGCGGTCTCC GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
Ugd29 CATGTCATCA GGCGGTCTCC GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
Ugd4 CATGTCATCA GGCGGTCTCC GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
Ugd12 CATGTCATCA GGCGGTCTGC GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
Ugd19 CATGTCATCA GGCGGTCTGC GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
Ugd2 CATGTCATCA GGCGGTCTGC GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
Ugd10 CATGTCATCA GGCGGTCTGC GCCACGGGAG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
BC-1 CATGTCATCA GGCGGTCTGT GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
Cons CATGTCATCA GGCGGTCT-- GCCACGGG-G AGCGAGCGTG TTGTATCCTT TGGCTACGGG
301 I/M N/H * * * A/G * * R/Q 360
BCBL-R CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGGAGTTTT GCTGCGAGAG
GK18 CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGGAGTTTT GCTGCGAGAG
Ugd15 CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGGAGTTTT GCTGCGAGAG
Ugd16 CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGGAGTTTT GCTGCGAGAG
Ugd23 CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGGAGTTTT GCTGCGAGAG
Ugd29 CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGGAGTTTT GCTGCGAGAG
Ugd4 CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGGAGTTTT GCTGCGAGAG
Ugd12 CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGCAGTTTT GCTGCGAGAG
Ugd19 CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGCAGTTTT GCTGCGAGAG
Ugd2 CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGCAGTTTT GCTGCGAGAG
Ugd10 CCAAACATAA ACCACAGGCC CACAACACTG TCCACAGAGC TTGCAGTGTT GCTGCGAGAG
BC-1 CCAAACATGC ACCACCGGCC CACAACATTG TCAACAGAGC TTGCAGTTCT GCTGCAAGAG
Cons CCAAACAT-- ACCAC-GGCC CACAACA-TG TC-ACAGAGC TTG-AGT--T GCTGC-AGAG
361 * 420
BCBL-R CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
GK18 CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
Ugd15 CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
Ugd16 CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
Ugd23 CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
Ugd29 CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
Ugd4 CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
Ugd12 CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
Ugd19 CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
Ugd2 CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
Ugd10 CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
BC-1 CTAGGATTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
Cons CTAGGA-TGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
421 * 480
BCBL-R CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACTGAC
GK18 CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACTGAC
Ugd15 CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACTGAC
Ugd16 CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACTGAC
Ugd23 CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACTGAC
Ugd29 CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACTGAC
Ugd4 CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACTGAC
Ugd12 CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACCGAC
Ugd19 CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACCGAC
Ugd2 CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACCGAC
Ugd10 CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACCGAC
BC-1 CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACCGAC
Cons CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATAC-GAC

```
        481                                                         540
BCBL-R  CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
  GK18  CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
 Ugd15  CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
 Ugd16  CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
 Ugd23  CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
 Ugd29  CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
  Ugd4  CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
 Ugd12  CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
 Ugd19  CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
  Ugd2  CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
 Ugd10  CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
  BC-1  CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
  Cons  CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC

        541                                     578
BCBL-R  ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATATC
  GK18  ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATATC
 Ugd15  ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATATC
 Ugd16  ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATATC
 Ugd23  ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATATC
 Ugd29  ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATATC
  Ugd4  ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATATC
 Ugd12  ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATATC
 Ugd19  ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATATC
  Ugd2  ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATATC
 Ugd10  ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATATC
  BC-1  ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATATC
  Cons  ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATATC
```

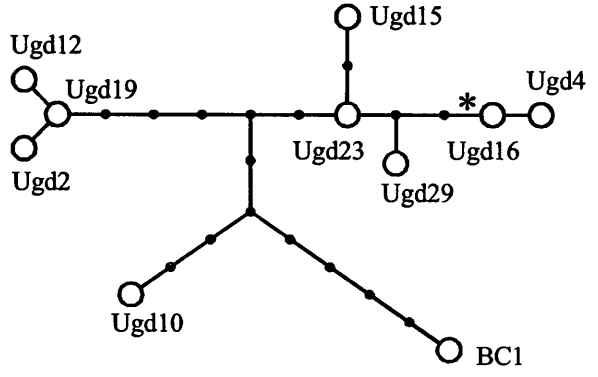
Fig. 6.3. Networks displaying relationships among DNA sequences in (A) K3-ORF26, (B) K9-T0.7/K12, (C) ORF75 and (D) K15 P loci.

Alignments were made consisting of sequences of the Ugandan strains and those of BC-1, BCBL-R and GK18. The length of the alignments and the number of substitution sites used to construct the networks are indicated. All substitution sites were used with the exception of one independently mutated site in T0.7/K12 (see text). Samples are represented by circles, whose sizes are proportional to the number of samples. Each subdivision represents a single substitution site; the asterisk indicates the single site in K9. Deletions are not represented. Rory Bowden helped to draw the networks in CorelDraw.

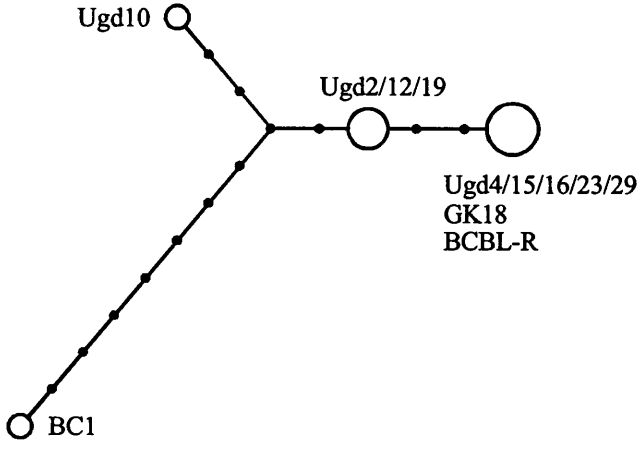
A K3 (519 bp; 10 sites) + ORF26 (495 bp; 6 sites)



B K9 (501 bp; 1 site *) + T0.7/K12 (551 bp; 24 sites)



C ORF75 (578 bp; 16 sites)



D K15 P (2086 bp; 34 sites)

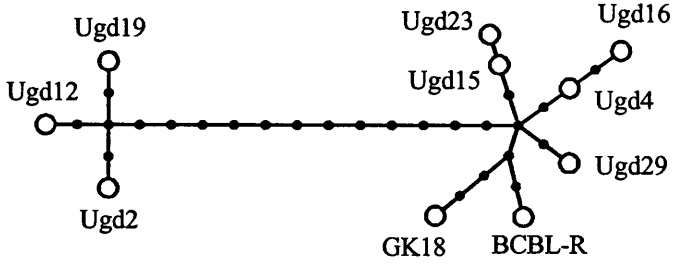


Table 6.1. Groups defined at each locus

Gene locus	Map Unit	No. of groups	Groups[a]
K1	0	3	A-(BC-1, Ugd4, Ugd12, Ugd16) B-(Ugd2, Ugd10, Ugd15, Ugd19, Ugd29) C-(Ugd23)
K3-ORF26	0.35	4	(BC-1) (Ugd2, Ugd12, Ugd19) (Ugd4, Ugd10, Ugd15, Ugd16, Ugd29) (Ugd23)
K9-T0.7/K12	0.85	3	(BC-1, Ugd10) (Ugd2, Ugd12, Ugd19) (Ugd4, Ugd15, Ugd16, Ugd23, Ugd29)
ORF75	0.97	3	(BC-1, Ugd10) (Ugd2, Ugd12, Ugd19) (Ugd4, Ugd15, Ugd16, Ugd23, Ugd29, GK18, BCBL-R)
K15	1.0	3	Same as for ORF75

[a] Strains that belong to the same group are enclosed in parentheses.

T0.7/K12, ORF75 and K15, suggesting recombination between ORF26 and K9-T0.7/K12. Thirdly, the K1 C variant (Ugd23) has a distinct sequence at K3-ORF26, but clusters with other Ugandan strains at the three remaining loci, suggesting recombination between ORF26 and K9-T0.7/K12.

There is no evidence for recombination in strains Ugd2, Ugd15, Ugd19 and Ugd29, as they co-segregate at all seven loci, Ugd2/19 in one group and Ugd15/29 in another group. Moreover, all strains co-segregate in K9-T0.7/K12, ORF75 and K15, thus providing no evidence for recombination within this region. Also, no evidence for recombination was noted within the loci examined.

Thus, network analysis, which is independent of previous categorisations of nucleotide pattern, indicates that five of the nine Ugandan strains (Ugd4, Ugd10, Ugd12, Ugd16 and Ugd23) are recombinants. It provides evidence for recombination between K1 and K3 and between ORF26 and T0.7/K12 in certain strains, but not between the three loci towards the right end of the genome.

6.4 VARIABILITY WITHIN THE LOCI

Variability among the Ugandan sequences at each locus was examined. Ten substitutions (nine synonymous and one non-synonymous) were identified in K3 (Fig. 6.2A), while six (three synonymous and three non-synonymous) were identified in ORF26 (Fig. 6.2B). Deletions were present in neither gene collection. Pairwise comparisons for K3 and ORF26 combined show a divergence among the sequences ranging from 0% (in identical sequences e.g. Ugd2/12/19; Fig. 6.3A) to 1% (between Ugd4/10 and Ugd2/12/19).

One synonymous substitution was identified in K9.

In T0.7/K12, all the sequences are different from each other (Fig. 6.3B). A total of 25 substitutions (24 shown in Fig. 6.3B) plus a unique 3-bp insertion in Ugd15 (Fig. 6.2D and 6.4B; see below) were identified. Seven (of which four are non-synonymous and three synonymous) are in the K12 gene and the rest are in the non-coding region (Fig. 6.2D). The synonymous substitution (position 90 in Fig. 6.2D) is not represented in the network (Fig. 6.3B) because it could not be resolved into a single mutation event. At this position, sequences Ugd10, Ugd12, Ugd16 and Ugd29 possess an A residue whereas all the other sequences possess a G residue. It is difficult to envisage a recombination event that could produce this polymorphism. Thus, the site most likely resulted from multiple independent mutations. This is the first time this observation has been made in T0.7/K12. All sequences published previously have a G residue at this position (Fig. 6.4B below).

Divergence in T0.7/K12 (shown in Table 6.2) ranges from 0.2% (Ugd4 vs. Ugd16) to 2.4% (BC-1 vs. Ugd12). That within the P-linked and within the two M-linked strains is 0.2-2.0% and 1.6%, respectively. Divergence between the M- and P-linked strains ranges from 1.5% (between Ugd10 and Ugd23) to 2.4% (between BC-1 and Ugd12).

Sixteeen substitutions (eight synonymous and eight non-synonymous) were identified in ORF75 (Fig. 6.2E and 6.3C). No deletions were present. Ugd10 and BC-1 each have eight and three unique mutations, respectively. Two groups of strains linked to the P allele are also evident at this locus, and the strains in each group are identical (Fig. 6.3C). Divergence between these two groups is minimal (0.5%; Table 6.2). The M-linked strains, on the other hand, show a considerably higher divergence (1.9%), almost four-fold compared to K15 P strains. Divergence between Ugd10 and the K15 P strains (0.9-1.4%) is less than that between Ugd10 and BC-1 (1.9%), indicating that Ugd10 is more closely related

Table 6.2. Pairwise divergence values[a] within and between groups in T0.7/K12, ORF75 and UPS75'

	T0.7/K12	ORF75	UPS75'
Size (bp)	548	578	209
Within Ugd2/12/19 (P)	0.2-0.4	0	0-0.5
Within Ugd4/15/16/23/29 (P)	0.4-1.1	0	0-0.5
Between Ugd2/12/19 (P) and Ugd4/15/16/23/29 (P)	1.1-2.0	0.5	0.5-1.4
Between Ugd10 (M) and BC-1 (M)	1.6	1.9	0
Between Ugd10 (M) and Ugd2/12/19 (P)	1.8-2.2	0.9	13.9-14.4
Between Ugd10 (M) and Ugd4/15/16/23/29 (P)	1.5-2.0	1.4	14.4-15.0
Between BC-1 (M) and Ugd2/12/19 (P)	2.0-2.4	1.7	13.9-14.4
Between BC-1 (M) and Ugd4/15/16/23/29 (P)	1.6-2.2	2.2	14.4-15.0

[a] Expressed as the percentage of substitutions per nucleotide. Ranges are given.

to the K15 P strains, in particular Ugd2/12/19 (from which it diverges by 0.9%), than to BC-1.

6.5 ANALYSIS OF ORF26, T0.7/K12 AND ORF75 SEQUENCES WITH PUBLISHED DATA

All the Ugandan sequences (30 for ORF26 and 10 each for T0.7/K12 and ORF75) were compared with sequences obtained from GenBank or published literature (Alagiozoglou et al., 2000; Poole et al., 1999; Zong et al., 1997). Where possible, all the groups identified previously were represented. This analysis enabled the Ugandan sequences to be categorised according to the genotypes established by Poole et al. (1999).

Sequences for ORF26, T0.7/K12, and some for ORF75 were generated from the listings of polymorphic sites and other information given by the authors (Poole et al., 1999; Zong et al., 1997). The alignments are shown in Fig. 6.4 (A, B, C). ORF75 sequences from South Africa were available from GenBank (Alagiozoglou et al., 2000). This alignment is shown in Fig. 6.4D. Two separate analyses were performed for ORF75, one comprising sequences generated from data by Zong et al. (1997) (Fig. 6.4C) and one comprising South African sequences (Fig. 6.4D). In the former case, ORF75 was analysed in conjunction with the 209 bp region (designated UPS75') which extends upstream from ORF75 to the point where the M and P alleles become distinct (Fig. 6.4C; Fig. 5.3 and 5.5).

The summarized data showing only substitution sites for sequences in this study and representative published sequences are shown in Fig. 6.5.

Fig. 6.4. Alignments of (A) ORF26, (B) T0.7/K12, (C) ORF75-UPS75' and (D) ORF75 DNA sequences of Ugandan and published strains.

Sequences of Ugandan strains (bold-only the first lines) were compared with published sequences (Alagiozoglou et al., 2000; Glenn et al., 1999; Poole et al., 1999; Zong et al., 1997). BCBL-R sequence is identical to that of BC-1 in (A), as is Ugd12 and Ugd15 to Ugd2/19 and Ugd4/16/23/29/BCBL-R/GK18, respectively, in (D). The groups to which sequences were (previous studies) or are (this study) assigned are indicated to the right of the first 6 blocks of each alignment. (A, C) are in the same orientation as the genomic sequence, and (B, D) are in the opposite orientation. Coordinates in the BC-1 genome for each loci: A, 47193-47490; B, 118065-117518; C, 133864-134650; D, 134441-133866. Positions of nucleotide changes are shown in bold-red, with the deletion (in B) represented by dots. The start (C) and stop (B) codons are highlighted in bold-blue.

A - ORF26

```
          1                                                          60
BCBL-R    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
  Erla    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
 BCBL1    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
   BC2    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
Ugd23     GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
431KAP    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
Ugd30     GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
 ST2-3    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
  Ugd1    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
 Ugd12    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
 Ugd19    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
  Ugd2    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
 Ugd21    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
 Ugd26    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
  Ugd7    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
  Ugd8    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
  Ugd9    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
  OKS3    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
 Ugd10    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
 Ugd11    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
 Ugd14    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
 Ugd15    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
 Ugd16    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
 Ugd17    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
 Ugd18    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
 Ugd20    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
 Ugd22    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
 Ugd24    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
 Ugd27    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
 Ugd28    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
 Ugd29    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
  Ugd3    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
  Ugd4    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
  Ugd5    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
  Ugd6    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
 ASM72    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
  TKS1    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCATTTTTTC AGTGGGACAG
 Ugd13    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
 Ugd25    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
  ZKS6    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CCCTTTTTTC AGTGGGACAG
  Cons    GGATCCCTCT GACAACCTTC AGATAAAAAA CGTATATGCC CC-TTTTTTC AGTGGGACAG
```

A
C3
B1
B2
B3/C2
C1
C3'
B3/C2

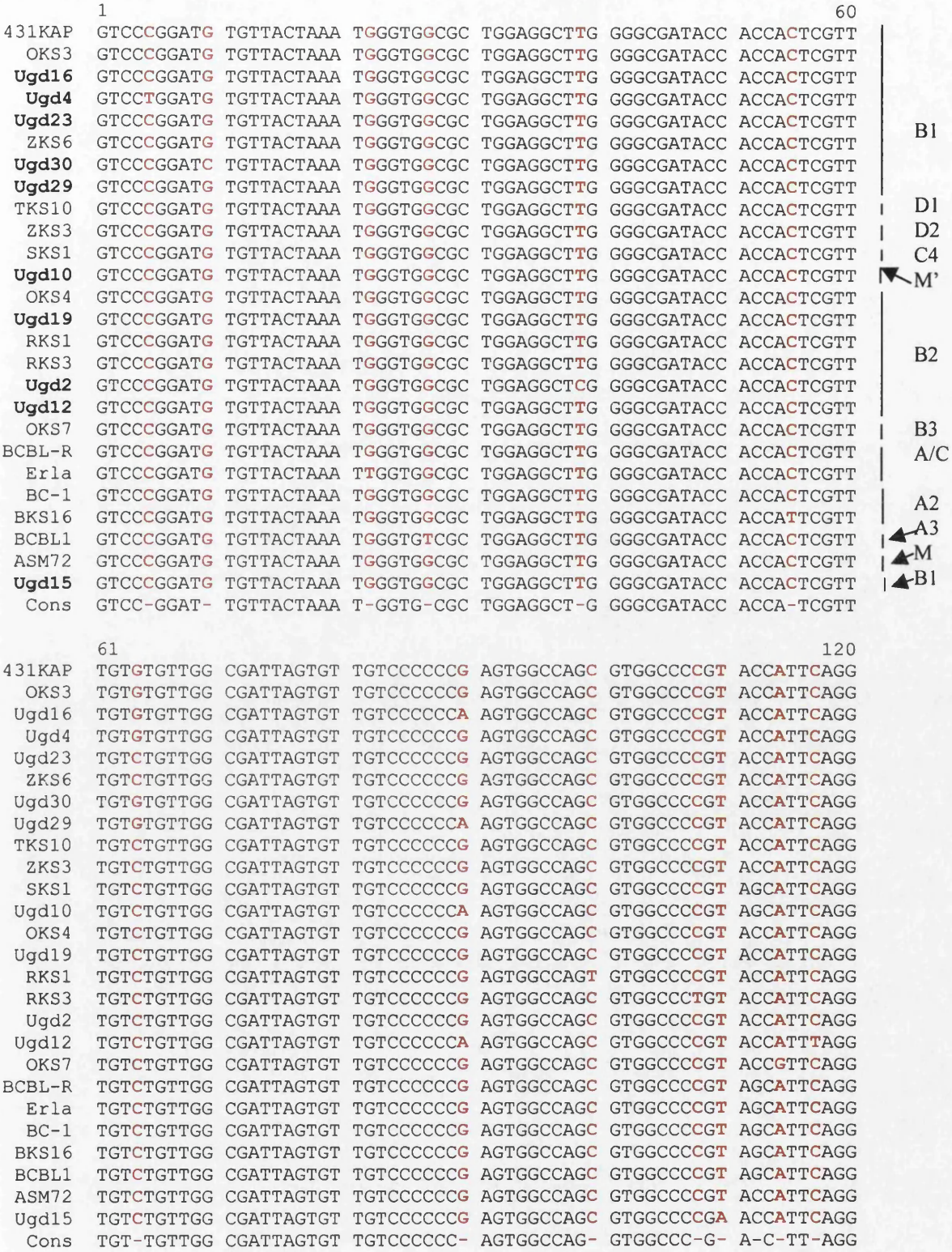
```
          61                                                        120
BCBL-R    CAACACCCAG CTAGCAGTGC TACCCCCATT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
  Erla    CAACACCCAG CTAGCAGTGC TACCCCCATT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
 BCBL1    CAACACCCAG CTAGCAGTGC TACCCCCATT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
   BC2    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
 Ugd23    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
431KAP    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
 Ugd30    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
 ST2-3    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
  Ugd1    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
 Ugd12    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
 Ugd19    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
  Ugd2    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
 Ugd21    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
 Ugd26    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
  Ugd7    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
  Ugd8    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
  Ugd9    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
  OKS3    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
 Ugd10    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
 Ugd11    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
 Ugd14    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
 Ugd15    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
 Ugd16    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
 Ugd17    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
 Ugd18    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
 Ugd20    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
 Ugd22    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
 Ugd24    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
 Ugd27    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
 Ugd28    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
 Ugd29    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
  Ugd3    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
  Ugd4    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
  Ugd5    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
  Ugd6    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
 ASM72    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
  TKS1    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
 Ugd13    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
 Ugd25    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAAGATTCCA CCATTGTGCT
  ZKS6    CAACACCCAG CTAGCAGTGC TACCCCCACT TTTTAGCCGA AAGGATTCCA CCATTGTGCT
  Cons    CAACACCCAG CTAGCAGTGC TACCCCCA-T TTTTAGCCGA AA-GATTCCA CCATTGTGCT
```

```
        121                                                     180
BCBL-R  CGAATCCAAC GGATTTGACC TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
  Erla  CGAATCCAAC GGATTTGACC TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
 BCBL1  CGAATCCAAC GGATTTGACC TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
   BC2  CGAATCCAAC GGATTTGACA TCGTGTTCCC CATGGTCGTG CCTCAGCAAC TGGGGCACGC
 Ugd23  CGAATCCAAC GGATTTGACA TCGTGTTCCC CATGGTCGTG CCTCAGCAAC TGGGGCACGC
431KAP  CGAATCCAAC GGATTTGACC TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
 Ugd30  CGAATCCAAC GGATTTGACC TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
 ST2-3  CGAATCCAAC GGATTTGACA TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
  Ugd1  CGAATCCAAC GGATTTGACA TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
 Ugd12  CGAATCCAAC GGATTTGACA TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
 Ugd19  CGAATCCAAC GGATTTGACA TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
  Ugd2  CGAATCCAAC GGATTTGACA TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
 Ugd21  CGAATCCAAC GGATTTGACA TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
 Ugd26  CGAATCCAAC GGATTTGACA TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
  Ugd7  CGAATCCAAC GGATTTGACA TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
  Ugd8  CGAATCCAAC GGATTTGACA TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
  Ugd9  CGAATCCAAC GGATTTGACA TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
  OKS3  CGAATCCAAC GGATTTGACC TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
 Ugd10  CGAATCCAAC GGATTTGACC TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
 Ugd11  CGAATCCAAC GGATTTGACC TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
 Ugd14  CGAATCCAAC GGATTTGACC TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
 Ugd15  CGAATCCAAC GGATTTGACC TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
 Ugd16  CGAATCCAAC GGATTTGACC TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
 Ugd17  CGAATCCAAC GGATTTGACC TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
 Ugd18  CGAATCCAAC GGATTTGACC TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
 Ugd20  CGAATCCAAC GGATTTGACC TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
 Ugd22  CGAATCCAAC GGATTTGACC TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
 Ugd24  CGAATCCAAC GGATTTGACC TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
 Ugd27  CGAATCCAAC GGATTTGACC TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
 Ugd28  CGAATCCAAC GGATTTGACC TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
 Ugd29  CGAATCCAAC GGATTTGACC TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
  Ugd3  CGAATCCAAC GGATTTGACC TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
  Ugd4  CGAATCCAAC GGATTTGACC TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
  Ugd5  CGAATCCAAC GGATTTGACC TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
  Ugd6  CGAATCCAAC GGATTTGACC TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
 ASM72  CGAATCCAAC GGATTTGACC TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
  TKS1  CGAATCCAAC GGATTTGACC TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
 Ugd13  CGAATCCAAC GGATTTGACC TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
 Ugd25  CGAATCCAAC GGATTTGACC TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
  ZKS6  CGAATCCAAC GGATTTGACC TCGTGTTCCC CATGGTCGTG CCGCAGCAAC TGGGGCACGC
  Cons  CGAATCCAAC GGATTTGAC- TCGTGTTCCC CATGGTCGTG CC-CAGCAAC TGGGGCACGC
```

```
        181                                                     240
BCBL-R  TATTCTGCAG CAGCTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGA
  Erla  TATTCTGCAG CAGCTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGA
 BCBL1  TATTCTGCAG CAGCTGTTGG TGTACCACAT ATACTCCAAA ATATCGGCCG GGGCCCCGGA
   BC2  TATTCTGCAG CAGCTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGG
 Ugd23  TATTCTGCAG CAGCTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGG
431KAP  TATTCTGCAG CAGCTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGG
 Ugd30  TATTCTGCAG CAGCTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGG
 ST2-3  TATTCTGCAG CAGCTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGG
  Ugd1  TATTCTGCAG CAGCTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGG
 Ugd12  TATTCTGCAG CAGCTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGG
 Ugd19  TATTCTGCAG CAGCTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGG
  Ugd2  TATTCTGCAG CAGCTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGG
 Ugd21  TATTCTGCAG CAGCTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGG
 Ugd26  TATTCTGCAG CAGCTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGG
  Ugd7  TATTCTGCAG CAGCTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGG
  Ugd8  TATTCTGCAG CAGCTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGG
  Ugd9  TATTCTGCAG CAGCTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGG
  OKS3  TATTCTGCAG CAGTTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGA
 Ugd10  TATTCTGCAG CAGTTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGA
 Ugd11  TATTCTGCAG CAGTTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGA
 Ugd14  TATTCTGCAG CAGTTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGA
 Ugd15  TATTCTGCAG CAGTTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGA
 Ugd16  TATTCTGCAG CAGTTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGA
 Ugd17  TATTCTGCAG CAGTTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGA
 Ugd18  TATTCTGCAG CAGTTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGA
 Ugd20  TATTCTGCAG CAGTTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGA
 Ugd22  TATTCTGCAG CAGTTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGA
 Ugd24  TATTCTGCAG CAGTTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGA
 Ugd27  TATTCTGCAG CAGTTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGA
 Ugd28  TATTCTGCAG CAGTTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGA
 Ugd29  TATTCTGCAG CAGTTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGA
  Ugd3  TATTCTGCAG CAGTTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGA
  Ugd4  TATTCTGCAG CAGTTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGA
  Ugd5  TATTCTGCAG CAGTTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGA
  Ugd6  TATTCTGCAG CAGTTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGA
 ASM72  TATTCTGCAG CAGTTGTTGG TATACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGA
  TKS1  TATTCTGCAG CAGTTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGA
 Ugd13  TATTCTGCAG CAGTTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGA
 Ugd25  TATTCTGCAG CAGTTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCG GGGCCCCGGA
  ZKS6  TATTCTGCAG CAGTTGTTGG TGTACCACAT CTACTCCAAA ATATCGGCCT GGGCCCCGGA
  Cons  TATTCTGCAG CAG-TGTTGG T-TACCACAT -TACTCCAAA ATATCGGCC- GGGCCCCGG-
```

```
          241                                                        298
BCBL-R    TGATGTAAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
  Erla    TGATGTAAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
 BCBL1    TGATGTAAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
   BC2    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
 Ugd23    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
431KAP    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
 Ugd30    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
 ST2-3    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
  Ugd1    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
 Ugd12    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
 Ugd19    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
  Ugd2    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
 Ugd21    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
 Ugd26    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
  Ugd7    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
  Ugd8    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
  Ugd9    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
  OKS3    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
 Ugd10    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
 Ugd11    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
 Ugd14    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
 Ugd15    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
 Ugd16    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
 Ugd17    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
 Ugd18    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
 Ugd20    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
 Ugd22    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
 Ugd24    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
 Ugd27    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
 Ugd28    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
 Ugd29    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
  Ugd3    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
  Ugd4    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
  Ugd5    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
  Ugd6    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
 ASM72    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
  TKS1    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
 Ugd13    TGATGTCAGT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
 Ugd25    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
  ZKS6    TGATGTCAAT ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
  Cons    TGATGT-A-T ATGGCGGAAC TTGATCTATA TACCACCAAT GTGTCATTTA TGGGGCGC
```

B - T0.7/K12


1
60
431KAP GTCCCGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
OKS3 GTCCCGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
Ugd16 GTCCCGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
Ugd4 GTCCTGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
Ugd23 GTCCCGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
ZKS6 GTCCCGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
Ugd30 GTCCCGGATC TGTTACTAAA TGGGTGGCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
Ugd29 GTCCCGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
TKS10 GTCCCGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
ZKS3 GTCCCGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
SKS1 GTCCCGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
Ugd10 GTCCCGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
OKS4 GTCCCGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
Ugd19 GTCCCGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
RKS1 GTCCCGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
RKS3 GTCCCGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
Ugd2 GTCCCGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTCG GGGCGATACC ACCACTCGTT
Ugd12 GTCCCGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
OKS7 GTCCCGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
BCBL-R GTCCCGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
Erla GTCCCGGATG TGTTACTAAA TTGGTGGCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
BC-1 GTCCCGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
BKS16 GTCCCGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTTG GGGCGATACC ACCATTCGTT
BCBL1 GTCCCGGATG TGTTACTAAA TGGGTGTCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
ASM72 GTCCCGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
Ugd15 GTCCCGGATG TGTTACTAAA TGGGTGGCGC TGGAGGCTTG GGGCGATACC ACCACTCGTT
Cons GTCC-GGAT- TGTTACTAAA T-GGTG-CGC TGGAGGCT-G GGGCGATACC ACCA-TCGTT
B1
D1
D2
C4
M'
B2
B3
A/C
A2
A3
M
B1
61
120
431KAP TGTGTGTTGG CGATTAGTGT TGTCCCCCCG AGTGGCCAGC GTGGCCCCGT ACCATTCAGG
OKS3 TGTGTGTTGG CGATTAGTGT TGTCCCCCCG AGTGGCCAGC GTGGCCCCGT ACCATTCAGG
Ugd16 TGTGTGTTGG CGATTAGTGT TGTCCCCCCA AGTGGCCAGC GTGGCCCCGT ACCATTCAGG
Ugd4 TGTGTGTTGG CGATTAGTGT TGTCCCCCCG AGTGGCCAGC GTGGCCCCGT ACCATTCAGG
Ugd23 TGTCTGTTGG CGATTAGTGT TGTCCCCCCG AGTGGCCAGC GTGGCCCCGT ACCATTCAGG
ZKS6 TGTCTGTTGG CGATTAGTGT TGTCCCCCCG AGTGGCCAGC GTGGCCCCGT ACCATTCAGG
Ugd30 TGTGTGTTGG CGATTAGTGT TGTCCCCCCG AGTGGCCAGC GTGGCCCCGT ACCATTCAGG
Ugd29 TGTGTGTTGG CGATTAGTGT TGTCCCCCCA AGTGGCCAGC GTGGCCCCGT ACCATTCAGG
TKS10 TGTCTGTTGG CGATTAGTGT TGTCCCCCCG AGTGGCCAGC GTGGCCCCGT ACCATTCAGG
ZKS3 TGTCTGTTGG CGATTAGTGT TGTCCCCCCG AGTGGCCAGC GTGGCCCCGT ACCATTCAGG
SKS1 TGTCTGTTGG CGATTAGTGT TGTCCCCCCG AGTGGCCAGC GTGGCCCCGT AGCATTCAGG
Ugd10 TGTCTGTTGG CGATTAGTGT TGTCCCCCCA AGTGGCCAGC GTGGCCCCGT AGCATTCAGG
OKS4 TGTCTGTTGG CGATTAGTGT TGTCCCCCCG AGTGGCCAGC GTGGCCCCGT ACCATTCAGG
Ugd19 TGTCTGTTGG CGATTAGTGT TGTCCCCCCG AGTGGCCAGC GTGGCCCCGT ACCATTCAGG
RKS1 TGTCTGTTGG CGATTAGTGT TGTCCCCCCG AGTGGCCAGT GTGGCCCCGT ACCATTCAGG
RKS3 TGTCTGTTGG CGATTAGTGT TGTCCCCCCG AGTGGCCAGC GTGGCCCTGT ACCATTCAGG
Ugd2 TGTCTGTTGG CGATTAGTGT TGTCCCCCCG AGTGGCCAGC GTGGCCCCGT ACCATTCAGG
Ugd12 TGTCTGTTGG CGATTAGTGT TGTCCCCCCA AGTGGCCAGC GTGGCCCCGT ACCATTTAGG
OKS7 TGTCTGTTGG CGATTAGTGT TGTCCCCCCG AGTGGCCAGC GTGGCCCCGT ACCGTTCAGG
BCBL-R TGTCTGTTGG CGATTAGTGT TGTCCCCCCG AGTGGCCAGC GTGGCCCCGT AGCATTCAGG
Erla TGTCTGTTGG CGATTAGTGT TGTCCCCCCG AGTGGCCAGC GTGGCCCCGT AGCATTCAGG
BC-1 TGTCTGTTGG CGATTAGTGT TGTCCCCCCG AGTGGCCAGC GTGGCCCCGT AGCATTCAGG
BKS16 TGTCTGTTGG CGATTAGTGT TGTCCCCCCG AGTGGCCAGC GTGGCCCCGT AGCATTCAGG
BCBL1 TGTCTGTTGG CGATTAGTGT TGTCCCCCCG AGTGGCCAGC GTGGCCCCGT AGCATTCAGG
ASM72 TGTCTGTTGG CGATTAGTGT TGTCCCCCCG AGTGGCCAGC GTGGCCCCGT ACCATTCAGG
Ugd15 TGTCTGTTGG CGATTAGTGT TGTCCCCCCG AGTGGCCAGC GTGGCCCCGA ACCATTCAGG
Cons TGT-TGTTGG CGATTAGTGT TGTCCCCCC- AGTGGCCAG- GTGGCCC-G- A-C-TT-AGG

```
          121                                stop                                180
431KAP    ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG... CCCTCCCGAG GAGTGTCAGT
  OKS3    ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG... CCCTCCCGAG GAGTGTCAGT
 Ugd16    ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG... CCCTCCCGAG GAGTGTCAGT
  Ugd4    ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG... CCCTCCCGAG GAGTGTCAGT
 Ugd23    ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG... CCCTCCCGAG GAGTGTCAGT
  ZKS6    ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG... CCCTCCCGAG GAGTGTCAGT
 Ugd30    ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG... CCCTCCCGAG GAGTGTCAGT
 Ugd29    ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG... CCCTCTCGAG GAGTGTCAGT
 TKS10    ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG... CCCTCCCGAA GAGTGTCAGT
  ZKS3    ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG... CCCTCCCGAA GAGTGTCAGT
  SKS1    ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG... CCCTCCCGAA GAGTGTCAGT
 Ugd10    ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG... CCCTCCCGAA GAGTGTCAGT
  OKS4    ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG... CCCTCCCGAA GAGTGTCAGT
 Ugd19    ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG... CCCTCCCGAA GAGTGTCAGT
  RKS1    ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG... CCCTCCCGAA GAGTGTCAGT
  RKS3    ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG... CCCTCCCGAA GAGTGTCAGT
  Ugd2    ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG... CCCTCCCGAA GAGTGTCAGT
 Ugd12    ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG... CCCTCCCGAA GAGTGTCAGT
  OKS7    ACACGAGTTG CAACGGGCGA GCACTGAAGC TAGCGTG... CCCTCCCGAA GAGTGTCAGT
BCBL-R    ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG... CCCTCCCAAA GAGTGTCAGT
  Erla    ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG... CCCTCCCAAA GAGTGTCAGT
  BC-1    ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG... CCCTCCCAAA GAGTGTCAGT
 BKS16    ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG... CCCTCCCAAA GAGTGTCAGT
 BCBL1    ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTG... CCCTCCCAAA GAGTGTCAGT
 ASM72    ACACGAGTTG CAACGGGTGC GCACTGAAGC TAGCGCG... CCCTCCCGAA GAGTGTCAGT
 Ugd15    ACACGAGTTG CAACGGGCGC GCACTGAAGC TAGCGTGCCC CCCCCCCGAG GAGTGTCAGT
  Cons    ACACGAGTTG CAACGGG-G- GCACTGAAGC TAGCG-G--- CCC-C-C-A- GAGTGTCAGT

          181                                                               240
431KAP    AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATCAAA CACATACAAT
  OKS3    AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATCAAA CACATACAAT
 Ugd16    AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATCAAA CACATACAAT
  Ugd4    AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATCAAA CACATACAAT
 Ugd23    AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATCAAA CACATACAAT
  ZKS6    AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATCAAA CACATACAAT
 Ugd30    AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATCAAA CACATACAAT
 Ugd29    AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATCAAA CACATACAAT
 TKS10    AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATCAAA CACATACAAT
  ZKS3    AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATCAAA CACATACAAT
  SKS1    AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATCAAA CACATACAAT
 Ugd10    AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATCAAA CACATACAAT
  OKS4    AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATCAAA CACATACAAT
 Ugd19    AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATCAAA CACATACAAT
  RKS1    AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATCAAA CACATACAAT
  RKS3    AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATCAAA CACATACAAT
  Ugd2    AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATCAAA CACATACAAT
 Ugd12    AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATCAAA CACATACAAT
  OKS7    AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATCAAA CACATACAAT
BCBL-R    AAAATAAAAT ACAAAAGCAC AATCACGGTT GCACCAAGCA CAACATTAAA CACATACAAT
  Erla    AAAATAAAAT ACAAAAGCAC AATCACGGTT GCACCAAGCA CAACATTAAA CACATACAAT
  BC-1    AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATTAAA CACATACAAT
 BKS16    AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATTAAA CACATACAAT
 BCBL1    AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATTAAA CACATACAAT
 ASM72    AAAATGAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATCAAA CACATACAAT
 Ugd15    AAAATAAAAT ACAAAAACAC AATCACGGTT GCACCAAGCA CAACATCAAA CACATACAAT
  Cons    AAAAT-AAAT ACAAAA-CAC AATCACGGTT GCACCAAGCA CAACAT-AAA CACATACAAT
```

```
          241                                                                300
431KAP    GCTGAAGAGC AGGCGTATCG AGGGCATACC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
  OKS3    GCTGAAGAGC AGGCGTATCG AGGGCATACC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
 Ugd16    GCTGAAGAGC AGGCGTATCG AGGGCATACC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
  Ugd4    GCTGAAGAGC AGGCGTATCG AGGGCATACC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
 Ugd23    GCTGAAGAGC AGGCGTATCG AGGGCATACC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
  ZKS6    GCTGAAGAGC AGGCGTATCG AGGGCATACC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
 Ugd30    GCTGAAGAGC AGGCGTATCG AGGGCATACC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
 Ugd29    GCTGAAGAGC AGGCGTATCG AGGGCATACC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
 TKS10    GCTGAAGAGC AGGCATATCG AGGGCATACC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
  ZKS3    GCTGAAGAGC AGGCATATCG AGGGCATACC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
  SKS1    GCTGAAGAGC AGGCGTATCG AGGGCATACC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
 Ugd10    CCTGAAGAGC AGGCGTATCG AGGGCATACC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
  OKS4    GCTGAAGAGT AGGAGTATCG AGGGCATACC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
 Ugd19    GCTGAAGAGT AGGAGTATCG AGGGCATATC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
  RKS1    GCTGAAGAGT AGGAGTATCG AGGGCATATC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
  RKS3    GCTGAAGAGT AGGAGTATCG AGGGCATATC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
  Ugd2    GCTGAAGAGT AGGAGTATCG AGGGCATATC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
 Ugd12    GCTGAAGAGT AGGAGTATCG AGGGCATATC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
  OKS7    GCTGAAGAGC AGGAGTATCA AGGGCATACC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
BCBL-R    GCTGAAGAGC AGGCGTATCG AGGGCATACC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
  Erla    GCTGAAGAGC AGGCGTATCG AGGGCATACC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
  BC-1    GCTGAAGAGC AGGCGTATCG AGGGCATACC CGTGGCTCAT AACACAGTCA CAGTTCAGAT
 BKS16    GCTGAAGAGC AGGCGTATCG AGGGCATACC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
 BCBL1    GCTGAAGAGC AGGCGTATCG AGGGCATACC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
 ASM72    GCCGAAGAGC AGGCGTATCG AGGGCATACC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
 Ugd15    GCTGAAGAGC AGGCGTATCG AGGGCATACC CGTGGCTCAT AACACAGTCA CAGTTCAGAA
  Cons    -C-GAAGAG- AGG--TATC- AGGGCATA-C CGTGGCTCAT AACACAGTCA CAGTTCAGA-

          301                                                                360
431KAP    GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
  OKS3    GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
 Ugd16    GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
  Ugd4    GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
 Ugd23    GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
  ZKS6    GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
 Ugd30    GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
 Ugd29    GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
 TKS10    GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
  ZKS3    GGCCGGCCCG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
  SKS1    GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
 Ugd10    GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
  OKS4    GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
 Ugd19    GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
  RKS1    GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
  RKS3    GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
  Ugd2    GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
 Ugd12    GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
  OKS7    GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
BCBL-R    GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
  Erla    GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
  BC-1    GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
 BKS16    GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
 BCBL1    GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
 ASM72    GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
 Ugd15    GGCCGGCACG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
  Cons    GGCCGGC-CG CGGTGTCAAC CAGGCCACCA TTCCTCTCCG CATTAAAGCA CTCGGTGGGG
```

```
          361                                                                420
431KAP    GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
  OKS3    GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
 Ugd16    GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
  Ugd4    GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
 Ugd23    GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
  ZKS6    GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
 Ugd30    GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
 Ugd29    GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
 TKS10    GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
  ZKS3    GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
  SKS1    GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
 Ugd10    GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
  OKS4    GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
 Ugd19    GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
  RKS1    GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
  RKS3    GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
  Ugd2    GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
 Ugd12    GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
  OKS7    GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
BCBL-R    GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
  Erla    GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
  BC-1    GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
 BKS16    GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
 BCBL1    GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
 ASM72    GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
 Ugd15    GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC
  Cons    GAGGGTGCCC TGGTTGACAC AATGTGCCGC GCATCAACCA GCGCGACGGA AGTTGGCCTC

          421                                                                480
431KAP    CAGCAAAGCA CATTTTGCAC GCAAATGGTG TCCGCTCTCC CAAACCACAC GAATGGCACC
  OKS3    CAGCAAAGCA CATTTTGCAC GCAAATGGTG TCCGCTCTCC CAAACCACAC GAATGGCACC
 Ugd16    CAGCAAAGCA CATTTTGCAC GCAAATGGTG TCCGCTCTCC CAAACCACAC GAATGGCACC
  Ugd4    CAGCAAAGCA CATTTTGCAC GCAAATGGTG TCCGCTCTCC CAAACCACAC GAATGGCACC
 Ugd23    CAGCAAAGCA CATTTTGCAC GCAAATGGTG TCCGCTCTCC CAAACCACAC GAATGGCACC
  ZKS6    CAGCAAAGCA CATTTTGCAC GCAAATGGTG TCCGCTCTCC CAAACCACAC GAATGGCACC
 Ugd30    CAGCAAAGCA CATTTTGCAC GCAAATGGTG TCCGCTCTCC CAAACCACAC GAATGGCACC
 Ugd29    CAGCAAAGCA CATTTTGCAC GCAAATGGTG TCCGCTCTCC CAAACCACAC GAATGGCACC
 TKS10    CAGCAAAGCA CATTTTGCAC GCAAATGGCG TCCGCTCTCC CAAACCACAC GAATGGCACC
  ZKS3    CAGCAAAGCA CATTTTGCAC GCAAATGGCG TCCGCTCTCC CAAACCACAC GAATGGCACC
  SKS1    CAGCAAAGCA CATTTTGCAC GCAAATGGCG TCCGCTCTCC CAAACCACAC GGATGGCACC
 Ugd10    CAGCAAAGCA CATTTTGCAC GCAAATGGCG TCCGCTCTCC CAAACCACAC GGATGGCACC
  OKS4    CAGCAAAGCA TATTTTGCAC GCAAATGGTG TCCGCTCTCC CAAACCACAC GAATGGCACC
 Ugd19    CAGCAAAGCA TATTTTGCAC GCAAATGGTG TCCGCTCTCC CAAACCACAC GAATGGCACC
  RKS1    CAGCAAAGCA TATTTTGCAC GCAAATGGTG TCCGCTCTCC CAAACCACAC GAATGGCACC
  RKS3    CAGCAAAGCA TATTTTGCAC GCAAATGGTG TCCGCTCTCC CAAACCACAC GAATGGCACC
  Ugd2    CAGCAAAGCA TATTTTGCAC GCAAATGGTG TCCGCTCTCC CAAACCACAC GAATGGCACC
 Ugd12    CAGCAAAGCA TATTTTGCAC GCAAATGGTG TCCGCTCTCC CAAACCACAC GAATGGCACC
  OKS7    CAGCAAAGCA CATTTTGCAC GCAAATGGTG TCCGCTCTCC CAAACCACAC GAATGGTACC
BCBL-R    CAGCAAAGCA CATTTTGCAC GCAAATGGCG TCCGCTCTCC CAAACCACAC GAATGGTACC
  Erla    CAGCAAAGCA CATTTTGCAC GCAAATGGCG TCCGCTCTCC CAAACCACAC GAATGGTACC
  BC-1    CAGCAAAGCA CATTTTGCAC GCAAATGGCG TCCGCTCTCC CAAACCACAC GAATGGTACC
 BKS16    CAGCAAAGCA CATTTTGCAC GCAAATGGCG TCCGCTCTCC CAAACCACAC GAATGGTACC
 BCBL1    CAGCAAAGCA CATTTTGCAC GCAAATGGCG TCCGCTCTCC CAAACCACAC GAATGGCACC
 ASM72    CAGCAAAGCA CATTTTGCAC GCAAATGGCG TCCGCTCTCC CAAACCACAC GAATGGCACC
 Ugd15    CAGCAAAGCA CATTTTGCAC GCAAATGGTG TCCGCTCTCC CAAACCACAC GAATGGCACC
  Cons    CAGCAAAGCA -ATTTTGCAC GCAAATGG-G TCCGCTCTCC CAAACCACAC G-ATGG-ACC
```

```
         481                                                        540
431KAP   ATGGCAAAAA ACACTCCCCC ACAATGTATT GACCTTGTTT GTTAATGAAA GTGTGGCACT
  OKS3   ATGGCAAAAA ACACTCCCCC ACAATGTATT GACCTTGTTT GTTAATGAAA GTGTGGCATT
 Ugd16   ATGGCAAAAA ACACTCCCCC ACAATGTATT GACCTTGTTT GTTAATGAAA GTGTGGCATT
  Ugd4   ATGGCAAAAA ACACTCCCCC ACAATGTATT GACCTTGTTT GTTAATGAAA GTGTGGCATT
 Ugd23   ATGGCAAAAA ACACTCCCCC ACAATGTATT GACCTTGTTT GTTAATGAAA GTGTGGCACT
  ZKS6   ATGGCAAAAA ACACTCCCCC ACAATGTATT GACCTTGTTT GTTAATGAAA GTGTGGCACT
 Ugd30   ATGGCAAAAA ACACTCCCCC ACAATGTATT GACCTTGTTT GTTAATGAAA GTGTGGCACT
 Ugd29   ATGGCAAAAA ACACTCCCCC ACAATGTATT GACCTTGTTT GTTAATGAAA GTGTGGCACT
 TKS10   ATGGCAAAAA ACACTCCCCC ACAATATATT GACCTTGTTT GTTAATGAAA TTGTGGCACT
  ZKS3   ATGGCAAAAA ACACTCCCCC ACAATATATT GACCTTGTTT GTTAATGAAA TTGTGGCACT
  SKS1   ATGGCAAAAA ACACTCCCCC ACAATGTATT GACCTTGTTT GTTAATGAAA TTGTGGCACT
 Ugd10   ATGGCAAAAA ACCCTCCCCC ACAATGTATT GACCTTGTTT GTTAATGAAA TTGTGGCACT
  OKS4   ATGGCAAAAA ACACTCCCCC ACAATGTATT GACCTTGTTT GTTAATGAAA TTGTGGCACT
 Ugd19   ATGGCAAAAA ACACTCCCCC ACAATGTATT GACCTTGTTT GTTAATGAAA TTGTGGCACT
  RKS1   ATGGCAAAAA ACACTCCCCC ACAATGTATT GACCTTGTTT GTTAATGAAA TTGTGGCACT
  RKS3   ATGGCAAAAA ACACTCCCCC ACAATGTATT GACCTTGTTT GTTAATGAAA TTGTGGCACT
  Ugd2   ATGGCAAAAA ACACTCCCCC ACAATGTATT GACCTTGTTT GTTAATGAAA TTGTGGCACT
 Ugd12   ATGGCAAAAA ACACTCCCCC ACAATGTATT GACCTTGTTT GTTAATGAAA TTGTGGCACT
  OKS7   ATGGCAAAAA ACACTCCCCC ACAATGTATT GACCTTGTTT GTTAATGAAA TTGTGGCACT
BCBL-R   ATGGCAAAAA ACACTCCCTC AGAATGTATT GACCTTGTTT GTTAATGAAA TTGTGGCACT
  Erla   ATGGCAAAAA ACACTCCCTC AGAATGTATT GACCTTGTTT GTTAATGAAA TTGTGGCACT
  BC-1   ATGGCAAAAA ACACTCCCTC ACAATGTATT GACCTTGTTT GTTAATGAAA TTGTGGCACT
 BKS16   ATGGCAAAAA ACACTCCCTC ACAATGTATT GACCTTGTTT GTTAATGAAA TTGTGGCACT
 BCBL1   ATGGCAAAAA ACACTCCCTC ACAATGTATT GACCTTGTTT GTTAATGAAA TTGTGGCACT
 ASM72   ATGGCAAAAA ACACTCCCTC ACAATGTATT GACCTTGTTT GTTAATGAAA TTGTGGCACT
 Ugd15   ATGGCAAAAA ACACTCCCCC ACAATGTATT GACCTTGTTT GTTAATGAAA GTGTGGCACT
  Cons   ATGGCAAAAA AC-CTCCC-C A-AAT-TATT GACCTTGTTT GTTAATGAAA -TGTGGCA-T

         541      551
431KAP   GACTTCGGCA G
  OKS3   GACTTCGGCA G
 Ugd16   GACTTCGGCA G
  Ugd4   GACTTCGGCA G
 Ugd23   GACTTCGGCA G
  ZKS6   GACTTCGGCA G
 Ugd30   GACTTCGGCA G
 Ugd29   GACTTCGGCA G
 TKS10   GACTTCGGCA G
  ZKS3   GACTTCGGCA G
  SKS1   GACTTCGGCA G
 Ugd10   GACTTCGGCA G
  OKS4   GACTTCGGCA G
 Ugd19   GACTTCGGCA G
  RKS1   GACTTCGGCA G
  RKS3   GACTTCGGCA G
  Ugd2   GACTTCGGCA G
 Ugd12   GACTTCGGCA G
  OKS7   GACTTCGGCA G
BCBL-R   GACTTCGGCA G
  Erla   GACTTCGGCA G
  BC-1   GACTTCGGCA G
 BKS16   GACTTCGGCA G
 BCBL1   GACTTCGGCA G
 ASM72   GACTTCGGCA G
 Ugd15   GACTTCGGCA G
  Cons   GACTTCGGCA G
```

C - ORF75-UPS75'

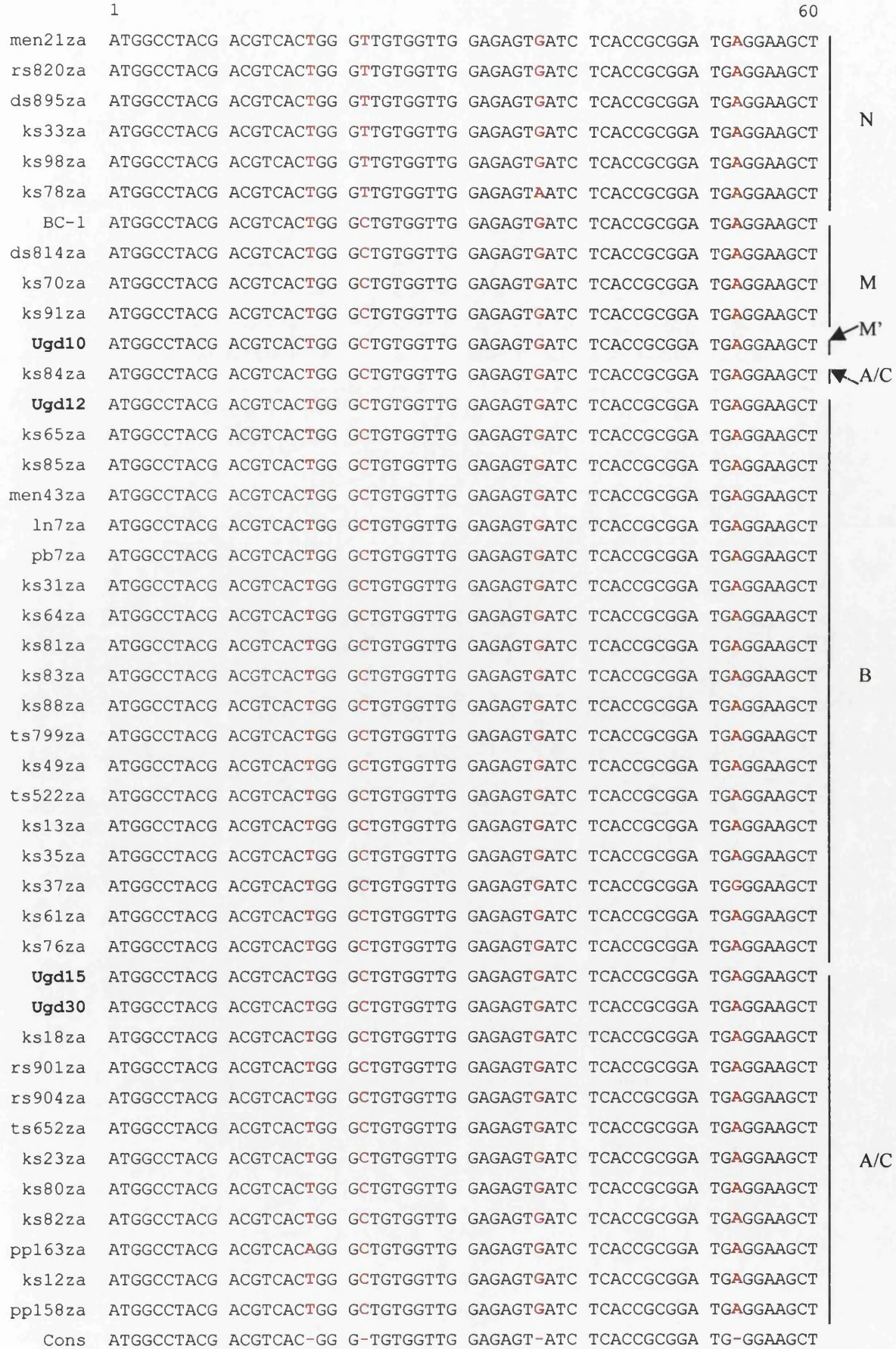
```
          181                                                        240
431KAP    CTGCCCACTT CCACTCTAGC CCACTCCTGC AGTCCTAGCT CTCGCAGCAA AACTGCAAGC
 Ugd12    CTGCCCACTT CCACTCTAGC CCACTCCTGC AGTCCTAGCT CTCGCAGCAA AACTGCAAGC
  Ugd2    CTGCCCACTT CCACTCTAGC CCACTCCTGC AGTCCTAGCT CTCGCAGCAA AACTGCAAGC
 Ugd19    CTGCCCACTT CCACTCTAGC CCACTCCTGC AGTCCTAGCT CTCGCAGCAA AACTGCAAGC
BCBL-R    CTGCCCACTT CCACTCTAGC CCACTCCTGC AGTCCTAGCT CTCGCAGCAA AACTCCAAGC
  GK18    CTGCCCACTT CCACTCTAGC CCACTCCTGC AGTCCTAGCT CTCGCAGCAA AACTCCAAGC
   ST1    CTGCCCACTT CCACTCTAGC CCACTCCTGC AGTCCTAGCT CTCGCAGCAA AACTCCAAGC
 Ugd15    CTGCCCACTT CCACTCTAGC CCACTCCTGC AGTCCTAGCT CTCGCAGCAA AACTCCAAGC
 Ugd23    CTGCCCACTT CCACTCTAGC CCACTCCTGC AGTCCTAGCT CTCGCAGCAA AACTCCAAGC
 Ugd29    CTGCCCACTT CCACTCTAGC CCACTCCTGC AGTCCTAGCT CTCGCAGCAA AACTCCAAGC
 Ugd16    CTGCCCACTT CCACTCTAGC CCACTCCTGC AGTCCTAGCT CTCGCAGCAA AACTCCAAGC
  Ugd4    CTGCCCACTT CCACTCTAGC CCACTCCTGC AGTCCTAGCT CTCGCAGCAA AACTCCAAGC
  C282    CTGCCCACTT CCACTCTAGC CCACTCCTGC AGTCCTAGCT CTCGCAGCAA AACTCCAAGC
  BC-1    CTGCCCACTT CCACTCTAGC CCACTCCTGC AATCCTAGCT CTTGCAGCAG AACTGCAAGC
 Ugd10    CTGCCCACTT CCACTCTAGC CCACTCCTGC AGTCCTAGCT CTCGCAGCAA CACTGCAAGC
  Cons    CTGCCCACTT CCACTCTAGC CCACTCCTGC A-TCCTAGCT CT-GCAGCA- -ACT-CAAGC

          241                                                        300
431KAP    TCTGTTGACA GTGTTGTGGG CCTGTGGTTT ATGTTTGGCC CGTAGCCAAA GGATACAACA
 Ugd12    TCTGTTGACA GTGTTGTGGG CCTGTGGTTT ATGTTTGGCC CGTAGCCAAA GGATACAACA
  Ugd2    TCTGTTGACA GTGTTGTGGG CCTGTGGTTT ATGTTTGGCC CGTAGCCAAA GGATACAACA
 Ugd19    TCTGTTGACA GTGTTGTGGG CCTGTGGTTT ATGTTTGGCC CGTAGCCAAA GGATACAACA
BCBL-R    TCTGTTGACA GTGTTGTGGG CCTGTGGTTT ATGTTTGGCC CGTAGCCAAA GGATACAACA
  GK18    TCTGTTGACA GTGTTGTGGG CCTGTGGTTT ATGTTTGGCC CGTAGCCAAA GGATACAACA
   ST1    TCTGTTGACA GTGTTGTGGG CCTGTGGTTT ATGTTTGGCC CGTAGCCAAA GGATACAACA
 Ugd15    TCTGTTGACA GTGTTGTGGG CCTGTGGTTT ATGTTTGGCC CGTAGCCAAA GGATACAACA
 Ugd23    TCTGTTGACA GTGTTGTGGG CCTGTGGTTT ATGTTTGGCC CGTAGCCAAA GGATACAACA
 Ugd29    TCTGTTGACA GTGTTGTGGG CCTGTGGTTT ATGTTTGGCC CGTAGCCAAA GGATACAACA
 Ugd16    TCTGTTGACA GTGTTGTGGG CCTGTGGTTT ATGTTTGGCC CGTAGCCAAA GGATACAACA
  Ugd4    TCTGTTGACA GTGTTGTGGG CCTGTGGTTT ATGTTTGGCC CGTAGCCAAA GGATACAACA
  C282    TCTGTTGACA GTGTTGTGGG CCTGTGGTTT ATGTTTGGCC CGTAGCCAAA GGATACAACA
  BC-1    TCTGTTGACA ATGTTGTGGG CCGGTGGTGC ATGTTTGGCC CGTAGCCAAA GGATACAACA
 Ugd10    TCTGTGGACA GTGTTGTGGG CCTGTGGTTT ATGTTTGGCC CGTAGCCAAA GGATACAACA
  Cons    TCTGT-GACA -TGTTGTGGG CC-GTGGT-- ATGTTTGGCC CGTAGCCAAA GGATACAACA

          301                                                        360
431KAP    CGCTCGCTCC CCCGTGGCGC AGACCGCCTG ATGACATGGG GATATCCAAG GAGCGGTGAC
 Ugd12    CGCTCGCTCC CCCGTGGCGC AGACCGCCTG ATGACATGGG GATATCCAAG GAGCGGTGAC
  Ugd2    CGCTCGCTCC CCCGTGGCGC AGACCGCCTG ATGACATGGG GATATCCAAG GAGCGGTGAC
 Ugd19    CGCTCGCTCC CCCGTGGCGC AGACCGCCTG ATGACATGGG GATATCCAAG GAGCGGTGAC
BCBL-R    CGCTCGCTCC CCCGTGGCGG AGACCGCCTG ATGACATGGG GATATCCAAG GAGCGGTGAC
  GK18    CGCTCGCTCC CCCGTGGCGG AGACCGCCTG ATGACATGGG GATATCCAAG GAGCGGTGAC
   ST1    CGCTCGCTCC CCCGTGGCGG AGACCGCCTG ATGACATGGG GATATCCAAG GAGCGGTGAC
 Ugd15    CGCTCGCTCC CCCGTGGCGG AGACCGCCTG ATGACATGGG GATATCCAAG GAGCGGTGAC
 Ugd23    CGCTCGCTCC CCCGTGGCGG AGACCGCCTG ATGACATGGG GATATCCAAG GAGCGGTGAC
 Ugd29    CGCTCGCTCC CCCGTGGCGG AGACCGCCTG ATGACATGGG GATATCCAAG GAGCGGTGAC
 Ugd16    CGCTCGCTCC CCCGTGGCGG AGACCGCCTG ATGACATGGG GATATCCAAG GAGCGGTGAC
  Ugd4    CGCTCGCTCC CCCGTGGCGG AGACCGCCTG ATGACATGGG GATATCCAAG GAGCGGTGAC
  C282    CGCTCGCTCC CCCGTGGCGG AGACCGCCTG ATGACATGGG GATATCCAAG GAGCGGTGAC
  BC-1    CGCTCGCTCC CCCGTGGCAC AGACCGCCTG ATGACATGGG GATATCCAAG GAGCGGTGAC
 Ugd10    CGCTCGCTCT CCCGTGGCGC AGACCGCCTG ATGACATGGG GATATCCAAG GAGCGGTGAC
  Cons    CGCTCGCTC- CCCGTGGC-- AGACCGCCTG ATGACATGGG GATATCCAAG GAGCGGTGAC
```

```
          361                                                        420
431KAP    AGCACAGCGA GCACCGTCTG TATTTCCACA TCCCGTCTCT CTCGCTCCTC CCTCGAAGTG
Ugd12     AGCACAGCGA GCACCGTCTG TATTTCCACA TCCCGTCTCT CTCGCTCCTC CCTCGAAGTG
Ugd2      AGCACAGCGA GCACCGTCTG TATTTCCACA TCCCGTCTCT CTCGCTCCTC CCTCGAAGTG
Ugd19     AGCACAGCGA GCACCGTCTG TATTTCCACA TCCCGTCTCT CTCGCTCCTC CCTCGAAGTG
BCBL-R    AGCACAGCGA GCACCGTCTG TATTTCCACA TCCCGTCTCT CTCGCTCCTC CCTCGAAGTG
GK18      AGCACAGCGA GCACCGTCTG TATTTCCACA TCCCGTCTCT CTCGCTCCTC CCTCGAAGTG
ST1       AGCACAGCGA GCACCGTCTG TATTTCCACA TCCCGTCTCT CTCGCTCCTC CCTCGAAGTG
Ugd15     AGCACAGCGA GCACCGTCTG TATTTCCACA TCCCGTCTCT CTCGCTCCTC CCTCGAAGTG
Ugd23     AGCACAGCGA GCACCGTCTG TATTTCCACA TCCCGTCTCT CTCGCTCCTC CCTCGAAGTG
Ugd29     AGCACAGCGA GCACCGTCTG TATTTCCACA TCCCGTCTCT CTCGCTCCTC CCTCGAAGTG
Ugd16     AGCACAGCGA GCACCGTCTG TATTTCCACA TCCCGTCTCT CTCGCTCCTC CCTCGAAGTG
Ugd4      AGCACAGCGA GCACCGTCTG TATTTCCACA TCCCGTCTCT CTCGCTCCTC CCTCGAAGTG
C282      AGCACAGCGA GCACCGTCTG TATTTCCACA TCCCGTCTCT CTCGCTCCTC CCTCGAAGTG
BC-1      AGCACAGCGA GCACCGTCTG TATTTCCACA TCCCGTCTCT CTCGCTCCTC CCTCGAAGTG
Ugd10     AGCACAGCGA GCACCGTCTG TATTTCCACA TCCCGTCTCT CTCGCTCCTC CCTCGAAGTG
Cons      AGCACAGCGA GCACCGTCTG TATTTCCACA TCCCGTCTCT CTCGCTCCTC CCTCGAAGTG

          421                                                        480
431KAP    GGAGGTCTTC GGAAAGTTAT CCATAGCAGA TAGTAGCCTC CGGTGCCACC GGGTACGAGA
Ugd12     GGAGGTCTTC GGAAAGTTAT CCATAGCAGA TAGTAGCCTC CGGTGCCACC GGGTACGAGA
Ugd2      GGAGGTCTTC GGAAAGTTAT CCATAGCAGA TAGTAGCCTC CGGTGCCACC GGGTACGAGA
Ugd19     GGAGGTCTTC GGAAAGTTAT CCATAGCAGA TAGTAGCCTC CGGTGCCACC GGGTACGAGA
BCBL-R    GGAGGTCTTC GGAAAGTTAT CCATAGCAGA TAGTAGCCTC CGGTGCCACC GGGTACGAGA
GK18      GGAGGTCTTC GGAAAGTTAT CCATAGCAGA TAGTAGCCTC CGGTGCCACC GGGTACGAGA
ST1       GGAGGTCTTC GGAAAGTTAT CCATAGCAGA TAGTAGCCTC CGGTGCCACC GGGTACGAGA
Ugd15     GGAGGTCTTC GGAAAGTTAT CCATAGCAGA TAGTAGCCTC CGGTGCCACC GGGTACGAGA
Ugd23     GGAGGTCTTC GGAAAGTTAT CCATAGCAGA TAGTAGCCTC CGGTGCCACC GGGTACGAGA
Ugd29     GGAGGTCTTC GGAAAGTTAT CCATAGCAGA TAGTAGCCTC CGGTGCCACC GGGTACGAGA
Ugd16     GGAGGTCTTC GGAAAGTTAT CCATAGCAGA TAGTAGCCTC CGGTGCCACC GGGTACGAGA
Ugd4      GGAGGTCTTC GGAAAGTTAT CCATAGCAGA TAGTAGCCTC CGGTGCCACC GGGTACGAGA
C282      GGAGGTCTTC GGAAAGTTAT CCATAGCAGA TAGTAGCCTC CGGTGCCACC GGGTACGAGA
BC-1      GGAGGTCTTC GGAAAGTTAT CCATAGCAGA TAGTAGCCTC CGGTGCCACC GGGTACGAGA
Ugd10     GGAGGTCTTC GGAAAGTTAT CCATAGCAGA TAGTAGCCTC CGGTGCCACC GGGTACGAGA
Cons      GGAGGTCTTC GGAAAGTTAT CCATAGCAGA TAGTAGCCTC CGGTGCCACC GGGTACGAGA

          481                                                        540
431KAP    GTGAGTGTGC CCGTACGGCT TGTATAAAAG CTGACAAAAG CTTCCTCATC CGCGGTGAGA
Ugd12     GTGAGTGTGC CCGTACGGCT TGTATAAAAG CTGACAAAAG CTTCCTCATC CGCGGTGAGA
Ugd2      GTGAGTGTGC CCGTACGGCT TGTATAAAAG CTGACAAAAG CTTCCTCATC CGCGGTGAGA
Ugd19     GTGAGTGTGC CCGTACGGCT TGTATAAAAG CTGACAAAAG CTTCCTCATC CGCGGTGAGA
BCBL-R    GTGAGTGTGC CCGTACGGCT TGTATAAAAG CTGACAAAAG CTTCCTCATC CGCGGTGAGA
GK18      GTGAGTGTGC CCGTACGGCT TGTATAAAAG CTGACAAAAG CTTCCTCATC CGCGGTGAGA
ST1       GTGAGTGTGC CCGTACGGCT TGTATAAAAG CTGACAAAAG CTTCCTCATC CGCGGTGAGA
Ugd15     GTGAGTGTGC CCGTACGGCT TGTATAAAAG CTGACAAAAG CTTCCTCATC CGCGGTGAGA
Ugd23     GTGAGTGTGC CCGTACGGCT TGTATAAAAG CTGACAAAAG CTTCCTCATC CGCGGTGAGA
Ugd29     GTGAGTGTGC CCGTACGGCT TGTATAAAAG CTGACAAAAG CTTCCTCATC CGCGGTGAGA
Ugd16     GTGAGTGTGC CCGTACGGCT TGTATAAAAG CTGACAAAAG CTTCCTCATC CGCGGTGAGA
Ugd4      GTGAGTGTGC CCGTACGGCT TGTATAAAAG CTGACAAAAG CTTCCTCATC CGCGGTGAGA
C282      GTGAGTGTGC CCGTACGGCT TGTATAAAAG CTGACAAAAG CTTCCTCATC CGCGGTGAGA
BC-1      GTGAGTGTGC CCGTACGGCT TGTATAAAAG TTCACAAAAG CTTCCTCATC CGCGGTGAGA
Ugd10     GTGAGTGTGC CCGTACGGCT TGTATAAAAG TTCACAAAAG CTTCCTCATC CGCGGTGAGA
Cons      GTGAGTGTGC CCGTACGGCT TGTATAAAAG -T-ACAAAAG CTTCCTCATC CGCGGTGAGA
```

```
          541                                    <start                    600
431KAP   TCACTCTCCA ACCACAGCCC AGTGACGTCG TAGGCCATGC CTAGAGGGCG CCCCGCCCCC
 Ugd12   TCACTCTCCA ACCACAGCCC AGTGACGTCG TAGGCCATGC CTAGAGGGCG CCCCGCCCCC
  Ugd2   TCACTCTCCA ACCACAGCCC AGTGACGTCG TAGGCCATGC CTAGAGGGCG CCCCGCCCCC
 Ugd19   TCACTCTCCA ACCACAGCCC AGTGACGTCG TAGGCCATGC CTAGAGGGCG CCCCGCCCCC
BCBL-R   TCACTCTCCA ACCACAGCCC AGTGACGTCG TAGGCCATGC CTAGAGGGTG CCCCGCCCCC
  GK18   TCACTCTCCA ACCACAGCCC AGTGACGTCG TAGGCCATGC CTAGAGGGTG CCCCGCCCCC
    ST1  TCACTCTCCA ACCACAGCCC AGTGACGTCG TAGGCCATGC CTAGAGGGTG CCCCGCCCCC
 Ugd15   TCACTCTCCA ACCACAGCCC AGTGACGTCG TAGGCCATGC CTAGAGGGTG CCCCGCCCCC
 Ugd23   TCACTCTCCA ACCACAGCCC AGTGACGTCG TAGGCCATGC CTAGAGGGTG CCCCGCCCCC
 Ugd29   TCACTCTCCA ACCACAGCCC AGTGACGTCG TAGGCCATGC CTAGAGGGTG CCCCGCCCCC
 Ugd16   TCACTCTCCA ACCACAGCCC AGTGACGTCG TAGGCCATGC CTAGAGGGTG CCCCGCCCCC
  Ugd4   TCACTCTCCA ACCACAGCCC AGTGACGTCG TAGGCCATGC CTAGAGGGTG CCCCGCCCCC
  C282   TCACTCTCCA ACCACAGCCC AGTGACGTCG TAGGCCATGC CTAGAGGGTG CCCCGCCCCC
  BC-1   TCACTCTCCA ACCACAGCCC AGTGACGTCG TAGGCCATGC CTAGAGGGCG CACCGCCCCC
 Ugd10   TCACTCTCCA ACCACAGCCC AGTGACGTCG TAGGCCATGC CTAGAGGGCG CACCGCCCCC
  Cons   TCACTCTCCA ACCACAGCCC AGTGACGTCG TAGGCCATGC CTAGAGGG-G C-CCGCCCCC

          601                                                         660
431KAP   GGGGACACCC TCTGTAGTCA GACTTGGGTT AGGGCTTTGA TTTCTCTGGG GAGTAGGAAG
 Ugd12   GGGGACACCC TCTGTAGTCA GACTTGGGTT AGGGCTTTGA TTTCTCTGGG GAGTAGGAAG
  Ugd2   GGGGACACCC TCTGTAGTCA GACTTGGGTT AGGGCTTTGA TTTCTCTGGG GAGTAGGAAG
 Ugd19   GGGGACACCC TCTGTAGTCA GACTTGGGTT AGGGCTTTGA TTTCTCTGGG GAGTAGGAAG
BCBL-R   GGGGACACCC TCTGTAGTCA GACTTGGGTT AGGGCTTTGA TTTCTCTGGG GAGTAGGAAG
  GK18   GGGGACACCC TCTGTAGTCA GACTTGGGTT AGGGCTTTGA TTTCTCTGGG GAGTAGGAAG
    ST1  GGGGACACCC TCTGTAGTCA GACTTGGGTT AGGGCTTTGA TTTCTCTGGG GAGTAGGAAG
 Ugd15   GGGGACACCC TCTGTAGTCA GACTTGGGTT AGGGCTTTGA TTTCTCTGGG GAGTAGGAAG
 Ugd23   GGGGACACCC TCTGTAGTCA GACTTGGGTT AGGGCTTTGA TTTCTCTGGG GAGTAGGAAG
 Ugd29   GGGGACACCC TCTGTAGTCA GACTTGGGTT AGGGCTTTGA TTTCTCTGGG GAGTAGGAAG
 Ugd16   GGGGACACCC TCTGTAGTCA GACTTGGGTT AGGGCTTTGA TTTCTCTGGG GAGTAGGAAG
  Ugd4   GGGGACACCC TCTGTAGTCA GACTTGGGTT AGGGCTTTGA TTTCTCTGGG GAGTAGGAAG
  C282   GGGGACACCC TCTGTAGTCA GACTTGGGTT AGGGCTTTGA TTTCTCTGGG GAGTAGGAAG
  BC-1   GGGGACACCC TCTGTAGTCA GGCTGCCGAG AAACCCGCGA GATCTCTGGG GAGTAGGAAG
 Ugd10   GGGGACACCC TCTGTAGTCA GGCTGCCGAG AAACCCGCGA GATCTCTGGG GAGTAGGAAG
  Cons   GGGGACACCC TCTGTAGTCA G-CT---G-- A---C---GA --TCTCTGGG GAGTAGGAAG

          661                                                         720
431KAP   AAACTGAGAA TCCCCAAATA TTACGCAGGC ACAGGTTGCT CTGCAGAGTT TGTTTTCGCT
 Ugd12   AAACTGAGAA TCCCCAAATA TTACGCAGGC ACAGGTTGCT CTGCAGAGTT TGTTTTCGCT
  Ugd2   AAACTGAGAA TCCCCAAATA TTACGCAGGC ACAGGTTGCT CTGCAGAGTT TGTTTTCGCT
 Ugd19   AAACTGAGAA TCCCCAAATA TTACGCAGGC ACAGGTTGCT CTGTAGAGTT TGTTTTCGCT
BCBL-R   AAACTGAGAA TCCCCAAATA TTACGCAGGC ACAGGTTGCT CTGCAGAGTT TGTTTTCGCT
  GK18   AAACTGAGAA TCCCCAAATA TTACGCAGGC ACAGGTTGCT CTGCAGAGTT TGTTTTCGCT
    ST1  AAACTGAGAA TCCCCAAATA TTACGCAGGC ACAGGTTGCT CTGCAGAGTT TGTTTTCGCT
 Ugd15   AAACTGAGAA TCCCCAAATA TTACGCAGGC ACAGGTTGCT CTGCAGAGTT TGTTTTCGCT
 Ugd23   AAACTGAGAA TCCCCAAATA TTACGCAGGC ACAGGTTGCT CTGCAGAGTT TGTTTTCGCT
 Ugd29   AAACTGAGAA TCCCCAAATA TTACGCAGGC ACAGGTTGCT CTGCAGAGTT TGTTTTCGCT
 Ugd16   AAACTGAGAA TCCCCAAATA TTACGCAGGC ACAGGTTGCT CTGCAGAGTT TGTTTTCGCT
  Ugd4   AAACTGAGAA TCCCCAAATA TTACGCAGGC ACAGGTTGCT CTGCAGAGTT TGTTTTCGCT
  C282   AAACTGAGAA TCCCCAAATA TTACGCAGGC ACAGGTTGCT CTGCAGAGTT TGTTTTCGCT
  BC-1   AAACTTAGAA TCCCCAAATA TGTCGCAGTC ACAGGTTGTC GGGCAGAGTC TGTTTCCGCT
 Ugd10   AAACTTAGAA TCCCCAAATA TGTCGCAGTC ACAGGTTGTC GGGCAGAGTC TGTTTCCGCT
  Cons   AAACT-AGAA TCCCCAAATA T--CGCAG-C ACAGGTTG-- --G-AGAGT- TGTTT-CGCT
```

```
         721                                                        780
431KAP   TTCGTGGAAT CCACAGTTAC ATGTAGCCAT GTCACTAACC TCAAATACTC AAAAAAAGGT
 Ugd12   TTCGTGGAAT CCACAGTTAC ATGTAGCCAT GTCACTAACC TCAAATACTC AAAAAAAGGT
  Ugd2   TTCGTGGAAT CCACAGTTAC ATGTAGCCAT GTCACTAACC TCAAATACTC AAAAAAAGGT
 Ugd19   TTCGTGGAAT CCACAGTTAC ATGTAGCCAT GTCACTAACC TCAAATACTC AAAAAAAGGT
BCBL-R   TTCGTGGAAT CCACAGTTAC ATGTAGCCAT GTCACTAACC TCAAATACTC AAAAAAAGGT
  GK18   TTCGTGGAAT CCACAGTTAC ATGTAGCCAT GTCACTAACC TCAAATACTC AAAAAAAGGT
   ST1   TTCGTGGAAT CCACAGTTAC ATGTAGCCAT GTCACTAACC TCAAATACTC AAAAAAAGGT
 Ugd15   TTCGTGGAAT CCACAGTTAC ATGTAGCCAT GTCACTAACC TCAAATACTC AAAAAAAGGT
 Ugd23   TTCGTGGAAT CCACAGTTAC ATGTAGCCAT GTCACTAACC TCAAATACTC AAAAAAAGGT
 Ugd29   TTCGTGGAAT CCACAGTTAC ATGTAGCCAT GTCACTAACC TCAAATACTC AAAAAAAGGT
 Ugd16   TTCGTCGAAT CCACAGTTAC ATGTAGCCAT GTCACTAACC TCAAATACTC AAAAAAAGGT
  Ugd4   TTCGTCGAAT CCACAGTTAC ATGTAGCCAT GTCACTAACC TCAAATACTC AAAAAAAGGT
  C282   TTCGTGGAAT CCACAGTTAC ATGTAGCCAT GTCACTAACC TCAAATACTC AAAAAAAGGT
  BC-1   TTCATGGGAT CCACAGTTAC TTGTAGCCAT GTCACTAACC TCAAATACTC AAAAAAAGCT
 Ugd10   TTCATGGGAT CCACAGTTAC TTGTAGCCAT GTCACTAACC TCAAATACTC AAAAAAAGCT
  Cons   TTC-T-G-AT CCACAGTTAC -TGTAGCCAT GTCACTAACC TCAAATACTC AAAAAAAG-T

         781 787
431KAP   ATCGATG
 Ugd12   ATCGATG
  Ugd2   ATCGATG
 Ugd19   ATCGATG
BCBL-R   ATCGATG
  GK18   ATCGATG
   ST1   ATCGATG
```

D - ORF75


1
60
men21za ATGGCCTACG ACGTCACTGG GTTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
rs820za ATGGCCTACG ACGTCACTGG GTTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
ds895za ATGGCCTACG ACGTCACTGG GTTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
ks33za ATGGCCTACG ACGTCACTGG GTTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
ks98za ATGGCCTACG ACGTCACTGG GTTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
ks78za ATGGCCTACG ACGTCACTGG GTTGTGGTTG GAGAGTAATC TCACCGCGGA TGAGGAAGCT
BC-1 ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
ds814za ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
ks70za ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
ks91za ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
Ugd10 ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
ks84za ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
Ugd12 ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
ks65za ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
ks85za ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
men43za ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
ln7za ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
pb7za ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
ks31za ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
ks64za ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
ks81za ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
ks83za ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
ks88za ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
ts799za ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
ks49za ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
ts522za ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
ks13za ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
ks35za ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
ks37za ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGGGGAAGCT
ks61za ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
ks76za ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
Ugd15 ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
Ugd30 ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
ks18za ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
rs901za ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
rs904za ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
ts652za ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
ks23za ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
ks80za ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
ks82za ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
pp163za ATGGCCTACG ACGTCACAGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
ks12za ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
pp158za ATGGCCTACG ACGTCACTGG GCTGTGGTTG GAGAGTGATC TCACCGCGGA TGAGGAAGCT
Cons ATGGCCTACG ACGTCAC-GG G-TGTGGTTG GAGAGT-ATC TCACCGCGGA TG-GGAAGCT
N
M
M'
A/C
B
A/C

```
          61                                                      120
men21za   TTTGTCAACT TCTATACTGG CCGAACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
rs820za   TTTGTCAACT TCTATACTGG CCGAACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
ds895za   TTTGTCAACT TCTATACTGG CCGAACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
 ks33za   TTTGTCAACT TCTATACTGG CCGAACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
 ks98za   TTTGTCAACT TCTATACTGG CCGAACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
 ks78za   TTTGTCAACT TCTATACTGG CCGAACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
   BC-1   TTTGTGAACT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
ds814za   TTTGTGAACT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
 ks70za   TTTGTGAACT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
 ks91za   TTTGTGAACT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
  Ugd10   TTTGTGAACT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
 ks84za   TTTGTGAACT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
  Ugd12   TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
 ks65za   TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
 ks85za   TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
men43za   TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
  ln7za   TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
  pb7za   TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
 ks31za   TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
 ks64za   TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
 ks81za   TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
 ks83za   TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACCCACTC TCGTACCCGG TGGCACCGGA
 ks88za   TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
ts799za   TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
 ks49za   TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
ts522za   TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
 ks13za   TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
 ks35za   TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
 ks37za   TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
 ks61za   TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
 ks76za   TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
  Ugd15   TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
  Ugd30   TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
 ks18za   TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
rs901za   TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
rs904za   TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
ts652za   TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
 ks23za   TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
 ks80za   TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
 ks82za   TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
pp163za   TTTGTCAGCT TTTATACAAT CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
 ks12za   TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG TGGCACCGGA
pp158za   TTTGTCAGCT TTTATACAAG CCGTACGGGC ACACTCACTC TCGTACCCGG CGGCACCGGA
    Cons  TTTGT-A-CT T-TATAC--- CCG-ACGGGC ACAC-CACTC TCGTACCCGG -GGCACCGGA
```

```
          121                                                       180
men21za   GGCTACTATT TGCTATGGAT AACTTTCCGA AGACCTCCCA CCTCGAGGGA GGAGCGAGAG
rs820za   GGCTACTATT TGCTATGGAT AACTTTCCGA AGACCTCCCA CCTCGAGGGA GGAGCGAGAG
ds895za   GGCTACTATT TGCTATGGAT AACTTTCCGA AGACCTCCCA CCTCGAGGGA GGAGCGAGAG
 ks33za   GGCTACTATT TGCTATGGAT AACTTTCCGA AGACCTCCCA CCTCGAGGGA GGAGCGAGAG
 ks98za   GGCTACTATT TGCTATGGAT AACTTTCCGA AGACCTCCCA CCTCGAGGGA GGAGCGAGAG
 ks78za   GGCTACTATT TGCTATGGAT AACTTTCCGA AGACCTCCCA CCTCGAGGGA GGAGCGAGAG
   BC-1   GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
ds814za   GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
 ks70za   GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
 ks91za   GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
  Ugd10   GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
 ks84za   GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
  Ugd12   GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
 ks65za   GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
 ks85za   GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
men43za   GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
  ln7za   GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
  pb7za   GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
 ks31za   GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
 ks64za   GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
 ks81za   GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
 ks83za   GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
 ks88za   GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
ts799za   GGCTACTATC TGCTATGGAT AACCTTCCGA AGACCTCCCA CTTCGAGGGA GGGGCGAGAG
 ks49za   GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
ts522za   GGCTACTATC TGCTATGGAT AACTTTCCGG AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
 ks13za   GGATACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
 ks35za   GGATACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
 ks37za   GGATACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
 ks61za   GGATACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
 ks76za   GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
  Ugd15   GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
  Ugd30   GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
 ks18za   GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
rs901za   GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
rs904za   GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
ts652za   GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
 ks23za   GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
 ks80za   GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
 ks82za   GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
pp163za   GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
 ks12za   GGCTACTATC TGCTACGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
pp158za   GGCTACTATC TGCTATGGAT AACTTTCCGA AGACCTCCCA CTTCGAGGGA GGAGCGAGAG
    Cons   GG-TACTAT- TGCTA-GGAT AAC-TTCCG- AGACCTCCCA C-TCGAGGGA GG-GCGAGAG
```

```
          181                                                       240
men21za   AGACGAGATG TGGAAGTGCA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
rs820za   AGACGAGATG TGGAAGTGCA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
ds895za   AGACGAGATG TGGAAGTGCA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
ks33za    AGACGAGATG TGGAAGTGCA GGCGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
ks98za    AGACGAGATG TGGAAGTGCA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
ks78za    AGACGAGATG TGGAAGTGCA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
BC-1      AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
ds814za   AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
ks70za    AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
ks91za    AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
Ugd10     AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
ks84za    AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
Ugd12     AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
ks65za    AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
ks85za    AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
men43za   AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
ln7za     AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
pb7za     AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
ks31za    AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
ks64za    AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
ks81za    AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
ks83za    AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
ks88za    AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
ts799za   AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
ks49za    AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT TACCGCTCCT TGGATATCCC
ts522za   AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATTTCCC
ks13za    AGACGGGATG TGGAAATGCA GACGGTGCTC GCTGTGCTGT CACCACTCCT TGGATATCCC
ks35za    AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCACTCCT TGGATATCCC
ks37za    AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCACTCCT TGGATATCCC
ks61za    AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCACTCCT TGGATATCCC
ks76za    AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCACTCCT TGGATATCCC
Ugd15     AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
Ugd30     AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
ks18za    AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
rs901za   AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
rs904za   AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
ts652za   AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
ks23za    AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
ks80za    AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
ks82za    AGACGGGATG TGGAAATACA GACGGCGCTC GCTGTGTTGT CACCGCTCCT TGGATATCCC
pp163za   AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
ks12za    AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CACCGCTCCT TGGATATCCC
pp158za   AGACGGGATG TGGAAATACA GACGGTGCTC GCTGTGCTGT CGCCGCTCCT TGGATATCCC
Cons      AGACG-GATG TGGAA-T-CA G-CGG-GCTC GCTGTG-TGT --CC-CTCCT TGGAT-TCCC
```

```
          241                                                        300
men21za   CATGTGATCA GGCGGTCCGC GCCACGGGGG AGCGAGCATG TTGTATCCTT TGGGTACGGG
rs820za   CATGTGATCA GGCGGTCCGC GCCACGGGGG AGCGAGCATG TTGTATCCTT TGGGTACGGG
ds895za   CATGTGATCA GGCGGTCCGC GCCACGGGGG AGCGAGCATG TTGTATCCTT TGGGTACGGG
 ks33za   CATGTGATCA GGCGGTCCGC GCCACGGGGG AGCGAGCATG TTGTATCCTT TGGGTACGGG
 ks98za   CATGTGATCA GGCGGTCCGC GCCACGGGGG AGCGAGCATG TTGTATCCTT TGGGTACGGG
 ks78za   CATGTGATCA GGCGGTCCGC GCCACGGGGG AGCGAGCATG TTGTATCCTT TGGGTACGGG
   BC-1   CATGTCATCA GGCGGTCTGT GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
ds814za   CATGTCATCA GGCGGTCTGT GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
 ks70za   CATGTCATCA GGCGGTCTGT GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
 ks91za   CATGTCATCA GGCGGTCTGT GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
  Ugd10   CATGTCATCA GGCGGTCTGC GCCACGGGAG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
 ks84za   CATGTCATCA GGCGGTCTGC GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
  Ugd12   CATGTCATCA GGCGGTCTGC GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
 ks65za   CATGTCATCA GGCGGTCTGC GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
 ks85za   CATGTCATCA GGCGGTCTGC GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
men43za   CATGTCATCA GGCGGTCTGC GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
  ln7za   CATGTCATCA GGCGGTCTGC GCCACGGGGG AACGAGCGTG TTGTATCCTT TGGCTACGGG
  pb7za   CATGTCATCA GGCGGTCTGC GCCACGGGGG AACGAGCGTG TTGTATCCTT TGGCTACGGG
 ks31za   CATGTCATCA GGCGGTCTGC GCCACGGGGG AACGAGCGTG TTGTATCCTT TGGCTACGGG
 ks64za   CATGTCATCA GGCGGTCTGC GCCACGGGGG AACGAGCGTG TTGTATCCTT TGGCTACGGG
 ks81za   CATGTCATCA GGCGGTCTGC GCCACGGGGG AACGAGCGTG TTGTATCCTT TGGCTACGGG
 ks83za   CATGTCATCA GGCGGTCTGC GCCACGGGGG AACGAGCGTG TTGTATCCTT TGGCTACGGG
 ks88za   CATGTCATCA GGCGGTCTGC GCCACGGGGG AACGAGCGTG TTGTATCCTT TGGCTACGGG
ts799za   CATGTCATCA GGCGGTCTGC GCCACGGGGG AACGAGCGTG TTGTATCCTT TGGCTACGGG
 ks49za   CATGTCATCA GGCGGTCTGC GCCACGGGGG AGCGAGCGTG TTGTATCTTT TGGCTACGGG
ts522za   CATGTCATCA GGCGGTCTGC GCCACGGGGG AACGAACGTG TTGTATCCTT TGGCTACGGG
 ks13za   CATGTCATCA GGCGGTCTAC GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
 ks35za   CATGTCATCA GGCGGTCTAC GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
 ks37za   CATGTCATCA GGCGGTCTAC GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
 ks61za   CATGTCATCA GGCGGTCTGC GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
 ks76za   CATGTCATCA GGCGGTCTGC GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
  Ugd15   CATGTCATCA GGCGGTCTCC GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
  Ugd30   CATGTCATCA GGCGGTCTCC GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
 ks18za   CATGTCATCA GGCGGTCTCC GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
rs901za   CATGTCATCA GGCGGTCTCC GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
rs904za   CATGTCATCA GGCGGTCTCC GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
ts652za   CATGTCATCA GGCGGTCTCC GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
 ks23za   CATGTCATCA GGCGGTCTGC GCCACGGGGG AACGAGCGTG TTGTATCCTT TGGCTACGGG
 ks80za   CATGTCATCA GGCGGTCTCC GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
 ks82za   CATGTCATCA GGCGGTCTCC GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
pp163za   CATGTCATCA GGCGGTCTCC GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
 ks12za   CATGTCATCA GGCGGTCTGC GACACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
pp158za   CATGCCATCA GGCGGTCTCC GCCACGGGGG AGCGAGCGTG TTGTATCCTT TGGCTACGGG
   Cons   CATG--ATCA GGCGGTC--- G-CACGGG-G A-CGA-C-TG TTGTATC-TT TGG-TACGGG
```

```
          301                                                       360
men21za   CCAGACAGGG ACCGCAGGCC TACGACACTG TCAACAGAGC TTGCAGTTCT GCTGCGAGAA
rs820za   CCAGACAGGG ACCGCAGGCC TACGACACTG TCAACAGAGC TTGCAGTTCT GCTGCGAGAA
ds895za   CCAGACAGGG ACCGCAGGCC TACGACACTG TCAACAGAGC TCGCAGTTCT GCTGCGAGAA
ks33za    CCAGACAGGG ACCGCAGGCC TACGACACTG TCAACAGAGC TTGCAGTTCT GCTGCGAGAA
ks98za    CCAGACAGGG ACCGCAGGCC TACGACACTG TCAACAGAGC TTGCAGTTCT GCTGCGAGAA
ks78za    CCAGACAGGG ACCGCAGGCC TACGACACTG TCAACAGAGC TTGCAGTTCT GCTGCGAGAA
  BC-1    CCAAACATGC ACCACCGGCC CACAACATTG TCAACAGAGC TTGCAGTTCT GCTGCAAGAG
ds814za   CCAAACATGC ACCACCGGCC CACAACATTG TCAACAGAGC TTGCAGTTCT GCTGCAAGAG
ks70za    CCAAACGTGC ACCACAGGCC CACAACATTA TCAACAGAGC TTGCAGTTCT GCTGCAAGAG
ks91za    CCAAACCTTC ACCACAGGCC CACAACATTA TCAACAGAGC TTCGAGTTCT GCTGCAAGAG
 Ugd10    CCAAACATAA ACCACAGGCC CACAACACTG TCCACAGAGC TTGCAGTGTT GCTGCGAGAG
ks84za    CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGGAGTTTT GCTGCGAGAG
 Ugd12    CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGCAGTTTT GCTGCGAGAG
ks65za    CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGCAGTTTT GCTGCGAGAG
ks85za    CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGCAGTTTT GCTGCGAGAG
men43za   CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGCAGTTTT GCTGCGAGAG
 ln7za    CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGCAGTTTT GCTGCGAGAG
 pb7za    CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGCAGTTTT GCTGCGAGAG
ks31za    CCAAACATAA ACCACGGGCC CACAACACTG TCAACAGAGC TTGCAGTTTT GCTGCGAGAG
ks64za    CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGCAGTTTT GCTGCGAGAG
ks81za    CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TCGCAGTTTT GCTGCGAGAG
ks83za    CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGCAGTTTT GCTGCGAGAG
ks88za    CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGCAGTTTT GCTGCGAGAG
ts799za   CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGCAGTTTT GCTGCGAGAG
ks49za    CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGCAGTTTT GCTGCGAGAG
ts522za   CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGCAGTTTT GCTGCGAGAG
ks13za    CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGCAGTTTT GCTGCGAGAG
ks35za    CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGCAGTTTT GCTGCGAGAG
ks37za    CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGCAGTTTT GCTGCGAGAG
ks61za    CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGCAGTTTT GCTGCGAGAG
ks76za    CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGCAGTTTT GCTGCGAGAG
 Ugd15    CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGGAGTTTT GCTGCGAGAG
 Ugd30    CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGGAGTTTT GCTGCGAGAG
ks18za    CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGGAGTTTT GCTGCGAGAG
rs901za   CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGGAGTTTT GCTGCGAGAG
rs904za   CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGGAGTTTT GCTGCGAGAG
ts652za   CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGGAGTTTT GCTGCGAGAG
ks23za    CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGGAGTTTT GCTGCGAGAG
ks80za    CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGGAGTTTT GCTGCGAGAG
ks82za    CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGGAGTTTT GCTGCGAGAG
pp163za   CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGGAGTTTT GCTGCGAGAG
ks12za    CCAAACATAA ACCACAGGCC CACAACATTG TCAACAGAGC TTGGAGTTTT GCTGCGAGAG
pp158za   CCAAACATAA ACCACAGGCC CACAACACTG TCAACAGAGC TTGGAGTTTT GCTGCGAGAG
   Cons   CCA-AC---- ACC-C-GGCC -AC-ACA-T- TC-ACAGAGC T---AGT--T GCTGC-AGA-
```

```
          361                                                        420
men21za   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGTAGGC ACCTGGTATC AAAAATCACA
rs820za   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGTAGGC ACCTGGTATC AAAAATCACA
ds895za   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGTAGGC ACCTGGTATC AAAAATCACA
 ks33za   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGTAGGC ACCTGGTATC AAAAATCACA
 ks98za   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGTAGGC ACCTGGTATC AAAAATCACA
 ks78za   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGTAGGC ACCTGGTATC AAAAATCACA
   BC-1   CTAGGATTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
ds814za   CTAGGATTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
 ks70za   CTAGGATTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
 ks91za   CTAGGATTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
  Ugd10   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
 ks84za   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
  Ugd12   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
 ks65za   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
 ks85za   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
men43za   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
  ln7za   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
  pb7za   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
 ks31za   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
 ks64za   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
 ks81za   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
 ks83za   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
 ks88za   CTAGGACTGC AAGAGTGGGC TAGAGTGGCA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
ts799za   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
 ks49za   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
ts522za   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
 ks13za   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
 ks35za   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
 ks37za   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
 ks61za   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
 ks76za   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
  Ugd15   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
  Ugd30   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
 ks18za   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
rs901za   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
rs904za   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
ts652za   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
 ks23za   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
 ks80za   CTAGGATTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
 ks82za   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
pp163za   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
 ks12za   CTAGGACTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
pp158za   CTAGGGCTGC AGGAGTGGGC TAGAGTGGAA GTGGGCAGGC ACCTGGTGTC CAAAATCACA
   Cons   CTAGG--TGC A-GAGTGGGC TAGAGTGG-A GTGGG-AGGC ACCTGGT-TC -AAAATCACA
```

```
          421                                                       480
men21za   CAGACTCTGC TAGAACCACA CCCACCTCAG TTTATAAGGG CATTTACACA AAATACCGAC
rs820za   CAGACTCTGC TAGAACCACA CCCACCTCAG TTTATAAGGG CATTTACACA AAATACCGAC
ds895za   CAGACTCTGC TAGAACCACA CCCACCTCAG TTTATAAGGG CATTTACACA AAATACCGAC
 ks33za   CAGACTCTGC TAGAACCACA CCCACCTCAG TTTATAAGGG CATTTACACA AAATACCGAC
 ks98za   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATAAGGG CATTTACACA AAATACCGAC
 ks78za   CAGACTCTGC TAGAACCACA CCCACCTCAG TTTATAAGGG CATTTACACA AAATACCGAC
   BC-1   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACCGAC
ds814za   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACCGAC
 ks70za   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACCGAC
 ks91za   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACCGAC
  Ugd10   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACCGAC
 ks84za   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACCGAC
  Ugd12   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACCGAC
 ks65za   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACCGAC
 ks85za   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACCGAC
men43za   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACCGAC
  ln7za   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACCGAC
  pb7za   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACCGAC
 ks31za   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACCGAC
 ks64za   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACCGAC
 ks81za   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACCGAC
 ks83za   CAGACCTTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACCGAC
 ks88za   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACCGAC
ts799za   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACCGAC
 ks49za   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACCGAC
ts522za   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACCGAC
 ks13za   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACCGAC
 ks35za   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACCGAC
 ks37za   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACCGAC
 ks61za   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACCGAC
 ks76za   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCGGGG CATTTACACA AAATACCGAC
  Ugd15   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACTGAC
  Ugd30   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACTGAC
 ks18za   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACTGAC
rs901za   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACTGAC
rs904za   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACTGAC
ts652za   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACTGAC
 ks23za   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACTGAC
 ks80za   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACTGAC
 ks82za   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACTGAC
pp163za   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACTGAC
 ks12za   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACTGAC
pp158za   CAGACCCTGC TAGAACCACA CCCACCTCAG TTTATCAGGG CATTTACACA AAATACTGAC
   Cons   CAGAC--TGC TAGAACCACA CCCACCTCAG TTTAT--GGG CATTTACACA AAATAC-GAC
```

```
          481                                                       540
men21za   CTGGTACCGT ACGAGGGGTT GGACGCGCCC GAGGGTCCCC CGCCCGTGGC TAGGCCACAC
rs820za   CTGGTACCGT ACGAGGGGTT GGACGCGCCC GAGGGTCCCC CGCCCGTGGC TAGGCCACAC
ds895za   CTGGTACCGT ACGAGGGGTT GGACGCGCCC GAGGGTCCCC CGCCCGTGGC TAGGCCACAC
 ks33za   CTGGTACCGT ACGAGGGGTT GGACGCGCCC GAGGGTCCCC CGCCCGTGGC TAGGCCACAC
 ks98za   CTGGTACCGT ACGAGGGGTT GGACGCGCCC GAGGGTCCCC CGCCCGTGGC TAGGCCACAC
 ks78za   CTGGTACCGT ACGAGGGGTT GGACGCGCCC GAGGGTCCCC CGCCCGTGGC TAGGCCACAC
   BC-1   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
ds814za   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
 ks70za   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCTGGC TAGGCCACAC
 ks91za   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
  Ugd10   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
 ks84za   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
  Ugd12   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
 ks65za   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
 ks85za   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
men43za   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
  ln7za   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
  pb7za   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
 ks31za   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
 ks64za   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
 ks81za   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
 ks83za   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
 ks88za   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
ts799za   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
 ks49za   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
ts522za   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
 ks13za   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
 ks35za   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
 ks37za   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
 ks61za   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
 ks76za   CTAGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
  Ugd15   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
  Ugd30   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
 ks18za   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
rs901za   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
rs904za   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
ts652za   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
 ks23za   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
 ks80za   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCTCGTGGC TAGGCCACAC
 ks82za   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
pp163za   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
 ks12za   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
pp158za   CTGGTACCGT ACGAGGGGTT GGAAGTGCCC GAGGGTCCCC AGCCCGTGGC TAGGCCACAC
    Cons   CT-GTACCGT ACGAGGGGTT GGA-G-GCCC GAGGGTCCCC -GC-C-TGGC TAGGCCACAC
```

```
          541                                 576
men21za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATG
rs820za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATG
ds895za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATG
 ks33za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATG
 ks98za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATG
 ks78za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ACGATG
   BC-1   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
ds814za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
 ks70za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
 ks91za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
  Ugd10   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
 ks84za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
  Ugd12   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
 ks65za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
 ks85za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
men43za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
  ln7za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
  pb7za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
 ks31za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
 ks64za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT GTGATA
 ks81za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
 ks83za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
 ks88za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
ts799za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
 ks49za   ATTGAAGATG ATGKCATTAT GCAGGCTGTT ATGATA
ts522za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
 ks13za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
 ks35za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
 ks37za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
 ks61za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
 ks76za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
  Ugd15   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
  Ugd30   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
 ks18za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
rs901za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
rs904za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
ts652za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
 ks23za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
 ks80za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
 ks82za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
pp163za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
 ks12za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
pp158za   ATTGAAGATG ATGTCATTAT GCAGGCTGTT ATGATA
   Cons   ATTGAAGATG ATG-CATTAT GCAGGCTGTT --GAT-
```

Fig. 6.5. Nucleotide changes within (A) ORF26, (B) T0.7/K12 and (C) ORF75-UPS75' loci.

Sequences of Ugandan strains (bold) were compared with published sequences (Fig. 6.4) and results of this analysis are summarized in this figure, showing the groups to which the Ugandan strains belong. Only representative published sequences for the groups to which the Ugandan sequences belong are also shown. Group designations are those defined by Poole et al. (1999). K1 subtypes are indicated. The results are presented in the context of variation from BCBL-R sequences; hyphens and dots denote identical and deleted residues, respectively. ORF26 and ORF75/UPS75' are in the same orientation as the genomic sequence, and T0.7/K12 is in the opposite orientation. The number at the top of each site indicates position in the alignment. PG, groups published previously; DG, groups determined in this study; nk, unknown.

A ORF26 (298 bp)

Isolate	K1	89	103	140	163	194	202	240	247	249	PG	DG
BCBL-R	A1	T	G	C	G	C	G	A	A	A	A	
BC-1	A2	-	-	-	-	-	-	-	-	-	A	
BC2	C3	C	-	A	T	-	-	G	C	-	C3	
Ugd23	C	C	-	A	T	-	-	G	C	-		C3
ST2-3	B	C	-	A	-	-	-	G	C	-	B2	
Ugd1	B	C	-	A	-	-	-	G	C	-		B2
Ugd12	A5	C	-	A	-	-	-	G	C	-		B2
Ugd19	B	C	-	A	-	-	-	G	C	-		B2
Ugd2	B	C	-	A	-	-	-	G	C	-		B2
Ugd21	B	C	-	A	-	-	-	G	C	-		B2
Ugd26	B	C	-	A	-	-	-	G	C	-		B2
Ugd7	B	C	-	A	-	-	-	G	C	-		B2
Ugd8	nk	C	-	A	-	-	-	G	C	-		B2
Ugd9	nk	C	-	A	-	-	-	G	C	-		B2
431KAP	B	C	-	-	-	-	-	G	C	-	B1	
Ugd30	nk	C	-	-	-	-	-	G	C	-		B1
OKS3	A5	C	-	-	-	T	-	-	C	-	B3/C2	
Ugd10	B	C	-	-	-	T	-	-	C	-		B3/C2
Ugd11	nk	C	-	-	-	T	-	-	C	-		B3/C2
Ugd14	nk	C	-	-	-	T	-	-	C	-		B3/C2
Ugd15	B	C	-	-	-	T	-	-	C	-		B3/C2
Ugd16	A5	C	-	-	-	T	-	-	C	-		B3/C2
Ugd17	nk	C	-	-	-	T	-	-	C	-		B3/C2
Ugd18	A5	C	-	-	-	T	-	-	C	-		B3/C2
Ugd20	nk	C	-	-	-	T	-	-	C	-		B3/C2
Ugd22	nk	C	-	-	-	T	-	-	C	-		B3/C2
Ugd24	A5	C	-	-	-	T	-	-	C	-		B3/C2
Ugd27	nk	C	-	-	-	T	-	-	C	-		B3/C2
Ugd28	nk	C	-	-	-	T	-	-	C	-		B3/C2
Ugd29	B	C	-	-	-	T	-	-	C	-		B3/C2
Ugd3	B	C	-	-	-	T	-	-	C	-		B3/C2
Ugd4	A5	C	-	-	-	T	-	-	C	-		B3/C2
Ugd5	nk	C	-	-	-	T	-	-	C	-		B3/C2
Ugd6	nk	C	-	-	-	T	-	-	C	G		B3/C2
Ugd13	B	C	-	-	-	T	-	-	C	-		B3/C2
Ugd25	nk	C	A	-	-	T	-	-	C	-		B3/C2
ASM72	C1	C	-	-	-	T	A	-	C	-	C1	

B T0.7/K12 (551 bp)

Isolate	K1	5	10	39	64	90	110	112	117	137	156	158	159	160	164	166	168	170	186	197	227	241	243	250	254	269	300	431	449	472	477	493	499	502	531	539	PG	DG
BCBL-R	A1	C	G	T	C	G	T	G	C	C	T	.	.	.	T	C	A	A	A	G	T	G	T	C	C	C	A	C	C	A	T	A	T	G	T	C	A/C	
BC-1	A2	-	-	-	-	-	-	-	-	-	-	.	.	.	-	-	-	-	-	A	-	-	-	-	-	-	T	-	-	-	-	-	-	C	-	-	A2	
OKS4	B	-	-	-	-	-	-	C	-	-	-	.	.	.	-	-	G	-	-	A	C	-	-	T	A	-	-	T	T	-	C	-	C	C	-	-	B2	
Ugd2	BI	-	-	C	-	-	-	C	-	-	-	.	.	.	-	-	G	-	-	A	C	-	-	T	A	T	-	T	T	-	C	-	C	C	-	-		B2
Ugd12	A5	-	-	-	-	A	-	C	T	-	-	.	.	.	-	-	G	-	-	A	C	-	-	T	A	T	-	T	T	-	C	-	C	C	-	-		B2
Ugd19	BI	-	-	-	-	-	-	C	-	-	-	.	.	.	-	-	G	-	-	A	C	-	-	T	A	T	-	T	T	-	C	-	C	C	-	-		B2
Ugd15	BIII	-	-	-	-	-	A	C	-	-	-	C	C	C	C	-	G	G	-	A	C	-	-	-	-	-	-	-	T	-	C	-	C	C	G	-		B1
Ugd29	BII	-	-	-	G	A	-	C	-	-	-	.	.	.	-	T	G	G	-	A	C	-	-	-	-	-	-	-	T	-	C	-	C	C	G	-		B1
Ugd23	C	-	-	-	-	-	-	C	-	-	-	.	.	.	-	-	G	G	-	A	C	-	-	-	-	-	-	-	T	-	C	-	C	C	G	-		B1
Ugd30	nk	-	C	-	G	-	-	C	-	-	-	.	.	.	-	-	G	G	-	A	C	-	-	-	-	-	-	-	T	-	C	-	C	C	G	-		B1
Ugd4	A5	T	-	-	G	-	-	C	-	-	-	.	.	.	-	-	G	G	-	A	C	-	-	-	-	-	-	-	T	-	C	-	C	C	G	T		B1
Ugd16	A5	-	-	-	G	A	-	C	-	-	-	.	.	.	-	-	G	G	-	A	C	-	-	-	-	-	-	-	T	-	C	-	C	C	G	T		B1
OKS3	A5	-	-	-	G	-	-	C	-	-	-	.	.	.	-	-	G	G	-	A	C	-	-	-	-	-	-	-	T	-	C	-	C	C	G	T	B1	
431KAP	B	-	-	-	G	-	-	C	-	-	-	.	.	.	-	-	G	G	-	A	C	-	-	-	-	-	-	-	T	-	C	-	C	C	G	-	B1	
ASM72	C	-	-	-	-	-	-	C	-	T	C	.	.	.	-	-	G	-	G	A	C	-	C	-	-	-	-	-	-	-	C	-	-	C	-	-	M	
Ugd10	BII	-	-	-	-	A	-	-	-	-	-	.	.	.	-	-	G	-	-	A	C	C	-	-	-	-	-	-	-	G	C	C	C	C	-	-		M'

C ORF75 (578 bp) UPS75' (209 bp)

Isolate	K1	53	102	212	223	230	231	235	236	246	249	251	263	269	270	272	310	319	320	511	513	PS	DS	589	592	622	625	626	627	629	630	632	633	634	636	637	638	641	642	666	682	683	689	699	700	701	702	704	710	716	724	726	728	741	779	PG	DG
BCBL-R	A1	C	A	G	C	A	A	C	C	T	C	G	T	T	T	T	C	G	G	C	G	A/C		T	C	A	T	G	G	T	T	G	G	G	T	T	T	T	T	G	T	A	G	C	T	C	T	C	T	T	G	G	A	A	G	A/C	
GK18	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A/C		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A/C	
ST1	B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A/C		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A/C	
Ugd15	BIII	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		A/C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		A/C
Ugd30	nk	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		A/C	unknown																																						
Ugd29	BII	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		A/C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		A/C
Ugd23	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		A/C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		A/C
Ugd4	A5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		A/C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-		A/C
Ugd16	A5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		A/C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-		A/C
Ugd2	BI	-	G	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	C	-	-		B	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		B
Ugd12	A5	-	G	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	C	-	-		B	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		B
Ugd19	BI	-	G	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	C	-	-		B	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-		B
431KAP	B	-	G	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	C	-	-	B		C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	B	
KS84ZA	nk	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	T	C	A/C		unknown																																				
Ugd10	BII	-	G	-	-	-	C	G	-	G	-	-	-	-	-	-	T	-	C	T	C		M'	C	A	G	G	C	C	A	G	A	A	C	C	G	C	G	A	T	G	T	T	T	C	G	G	-	C	C	A	-	G	T	C		M
BC-1	A	-	G	A	T	G	-	G	-	-	-	A	G	G	C	-	-	A	C	T	C	M		C	A	G	G	C	C	A	G	A	A	C	C	G	C	G	A	T	G	T	T	T	C	G	G	-	C	C	A	-	G	T	C	M	
DS814ZA	nk	-	G	A	T	G	-	G	-	-	-	A	G	G	C	-	-	A	C	T	C	M		unknown																																				
KS70ZA	nk	G	G	A	T	G	-	G	-	-	T	A	-	G	C	C	-	A	C	T	C	M		unknown																																				
KS91ZA	nk	-	G	A	T	G	-	-	G	-	T	A	-	G	A	G	-	A	C	T	C	M		unknown																																				

Fig. 6.5A shows that all the Ugandan sequences clustered into patterns described previously in ORF26. The majority (19) have the B3/C2 pattern (like OKS3; Table 2.1), nine have the B2 pattern (like ST-3; Table 2.1), one (Ugd30) has the B1 pattern (like 431KAP; Table 2.1) and one (Ugd23) has the C3 pattern (like BC2, Table 2.1). Notably, none of the 30 Ugandan sequences grouped with the A subtypes.

Also in T0.7/K12, the Ugandan strains (with the exception of Ugd10) fall into nucleotide patterns reported previously (Fig. 6.5B). Six (Ugd4/15/16/23/29/30) cluster with strains of the B1 nucleotide pattern (e.g. 431KAP) and three (Ugd2/12/19) with strains of the B2 pattern (e.g. OKS3). The M pattern in T0.7/K12 is represented by ASM72 (Table 2.1). BC-1 belongs to pattern A2 at this locus. The pattern of Ugd10 is distinct, but it shares characteristics with both B and M patterns at positions 168, 197, 227, 477 and 502 (Fig. 6.4B). This pattern has therefore been denoted M'. As in ORF26, none of the Ugandan sequences clustered with sequences of the A (e.g. BC-1) or A/C (e.g. BCBL-R) patterns.

In ORF75 (Fig. 6.5C), the five sequences (Ugd4/15/16/23/29) that have a B1 pattern in T0.7/K12 are identical, but, surprisingly, cluster with sequences that have the pattern A/C (e.g. BCBL-R). The three Ugandan sequences (Ugd2/12/19) that have a B2 pattern at T0.7/K12 are also identical and cluster with sequences that have pattern B (e.g. 431KAP). The M pattern in ORF75 is represented by BC-1, but this sequence is identical to that of ASM72. Ugd10 shows a distinct pattern in ORF75 as well, and also shares characteristics with both the B and M patterns (positions 102, 235 and 320; Fig. 6.5C). This pattern has also been designated as M'. Interestingly, the ORF75 sequence of Ugd10 is closely related to that of KS84ZA (Fig. 6.4D and 6.5C; Table 2.1), a sequence categorised among the A/C group previously (Alagiozoglou et al., 2000). The M

sequences in samples from South Africa (DS814ZA, KS70ZA and KS91ZA; Table 2.1) are more closely related to BC-1 than to Ugd10 (Fig. 6.4D and 6.5C). It is clear from Fig 6.4D that the N subtype from South Africa (Alagiozoglou et al., 2000) is distinct from all other sequences in the alignment.

Analysis of the 209 bp region upstream from ORF75 (UPS75') revealed 32 substitution sites (Fig. 6.5C). Surprisingly, in this region Ugd10 is identical to BC-1. The sequences linked to the P allele show variation ranging from 0% (in identical sequences) to 1.4% (Ugd4/16 vs. Ugd19; Fig. 6.5C and Table 6.2). However, the M-linked strains differ from the P-linked strains by up to 15%.

In summary, comparison of the Ugandan sequences with those published previously revealed that in ORF75 Ugd4/15/16/23/29/30 have the same pattern as BCBL-R and all other sequences that have the same sequence as BCBL-R, while Ugd2/12/19 belong to a different group with 431KAP (Table 2.1) and B pattern sequences from South Africa (Fig. 6.4C, D). The fact that Ugd10 shares characteristics with the B and M groups in T0.7/K12 and ORF75 suggests that it is evolutionarily related to these two groups.

6.5.1 Overall genotypes of the Ugandan strains

Analysis of the patterns identified in ORF26, T0.7/K12 and ORF75 together with those of K1 and K15 resulted in the overall genotypes summarised in Table 6.3. The two K15 P groups have been designated as B and A/C because of the continuity of their respective patterns into the B and A/C patterns of the immediately adjacent portion of ORF75.

The grouping of Ugandan sequences based on nucleotide patterns also reveals evidence for recombination between K1 and ORF26, and between ORF26 and

Table 6.3. Subtype designations at each locus and overall genotypes

Sample[a]	K1	ORF26	T0.7/K12	ORF75	UPS75'	K15	Overall[b]	Revised Overall
Ugd4	A5	B3/C2	B1	A/C	A/C	P(A/C)	A5-B-A	A5-BII
Ugd16	A5	B3/C2	B1	A/C	A/C	P(A/C)	A5-B-A	A5-BII
Ugd12	A5	B2	B2	B	B	P(B)	A5-B	A5-BI
Ugd2	BI	B2	B2	B	B	P(B)	B	BI
Ugd19	BI	B2	B2	B	B	P(B)	B	BI
431KAP	BII[c]	B1	B1	B	B	P[d]	B	BII-BI
Ugd30	nk	B1	B1	A/C	nk	P[d]	nk-B-A/C	nk-BII
Ugd15	BIII	B3/C2	B1	A/C	A/C	P(A/C)	B-A/C	BIII-BII
Ugd29	BII	B3/C2	B1	A/C	A/C	P(A/C)	B-A/C	BII
Ugd23	C	C3	B1	A/C	A/C	P(A/C)	C-B-C	C-BII
BC2	C3	C3	A/C	A/C	A/C	P[d]	C	C-BII
BCBL-R	A1	A	A/C	A/C	A/C	P(A/C)	A	A-BII
Ugd10	BII	B3/C2	M'	M'	M	M(M')	B-M'	BII-M'
ASM72	C1	C1	M	M	nk	M[d]	C-M	C-M
BC-1	A2	A	A	M	M	M(M)	A-M	A-M

[a] Ugandan samples are in bold, and selected sequences (Table 2.4) used in the comparative analysis are also shown. Subtypes are designated according to the nomenclature of Poole *et al.* (1999), except for K15 and the revised overall genotypes that were designated on the basis of observations made in this study. The use of slashes (e.g. B3/C2) indicates that subtypes could not be distinguished.

[b] The most parsimonious solution is shown.

[c] 431KAP clustered with UKma24 (Lacoste et al., 2000a), therefore in this study it falls into BII with UKma24 (Fig. 4.6).

[d] K15 genotype based on PCR assay only, and not on DNA sequence (Poole et al., 1999; this study for Ugd30). In determining the overall genotype, the K15 genotype was assumed to be the same as that for ORF75.

Abbreviation: nk, unknown.

T0.7/K12 in these strains. Ugd2/19 and Ugd15/29 co-segregate at all six loci, suggesting that their genomes have not been subjected to detectable recombination. Except in K1, Ugd12 has the same type of genome as Ugd2/19, and Ugd4/16 the same type as Ugd15/29. Ugd23 has the same type of genome as Ugd4/15/16/29, except in K1 and ORF26. Ugd10 has the same genotype as Ugd29 in K1 and ORF26, but it is distinct in T0.7/K12 and ORF75, and is closely related to BC-1 in UPS75'-K15.

It is interesting to note that BCBL-R (which in ORF26, T0.7/K12 and ORF75 has the same sequence as several strains with K1 A or C genes) groups with Ugd4/15/16/23/29 in ORF75/UPS75' and K15 but not at other loci (with the exception of K1). This suggests recombination between T0.7/K12 and ORF75, which was not evident from the network analysis (section 6.3 above). Therefore, BCBL-R and its counterparts probably have recombinant genomes. Thus, although the group designations (Table 6.3) suggest that Ugd4/15/16/23/29 may be recombinants between T0.7/K12 and ORF75, this may not be the case. Furthermore, 431KAP (the strain regarded as the prototype for subtype B strains; Poole et al., 1999) has the pattern of Ugd2/12/19 in ORF75 and the pattern of Ugd4/15/16/23/29 in T0.7/K12, suggesting recombination between members of the same subtype at a point between T0.7/K12 and ORF75.

These observations highlight the problems associated with fixing prototype strains on the basis of "weight of numbers", as was done in previous studies (Poole et al., 1999; Zong et al., 1999). To gain further insights into the relationships of these strains, a network analysis was performed.

6.5.2 Network analysis of Ugandan sequences with selected published sequences

The networks (Fig. 6.3) for ORF26, T0.7/K12 and ORF75 were modified using data in Fig. 6.5 to include sequences of Ugd30 (which clustered with 431KAP in ORF26), 431KAP and BCBL-R. The modified networks are shown in Fig. 6.6.

Examination of the groups, together with those for K1 and K15 (Table 6.1), reveals that BCBL-R (which belongs to subtype A in K1) clusters with BC-1 in K1, ORF26 and T0.7/K12 (Fig. 6.6A,B), but clusters with Ugd4/15/16/23/29 in ORF75 (Fig. 6.6C) and K15 (Fig. 6.3D), suggesting recombination between T0.7/K12 and ORF75 in the BCBL-R lineage. 431KAP (which belongs to subtype B K1) clusters with Ugd4/15/16/23/29 in T0.7/K12 (Fig. 6.6B) but with Ugd2/12/19 in ORF75 (Fig. 6.6C), suggesting recombination between T0.7/K12 and ORF75 in this strain. Ugd30 clusters with 431KAP in ORF26 (Fig. 6.6A) and with Ugd4/15/16/23/29 in T0.7/K12 and ORF75 (Fig. 6.6B, C). Interestingly, BC-1 appears to maintain the same position throughout.

Thus, the overall genotypes may be revised as shown in the last column of Table 6.3. Ugd2/19 and Ugd29 (whose genomes show no evidence for recombination) are considered to represent two sublineages (BI and BII, respectively) within the B lineage. The designation has been based on the fact that K1 phylogenetic analysis indicated that both Ugd2/19 and Ugd29 lineages belong to subtype B. Ugd15 belongs to a distinct K1 B cluster (BIII; see Figure 4.6).

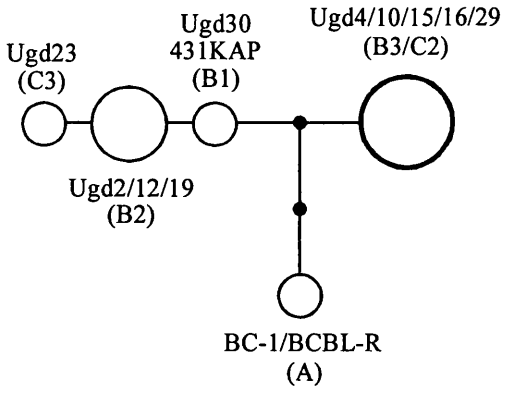
6.6 CONCLUSION AND DISCUSSION

Ugd4, Ugd10, Ugd12, Ugd16, Ugd23 and BCBL-R have genomes that may reflect a history of recombination between subtypes, while Ugd15 and 431KAP

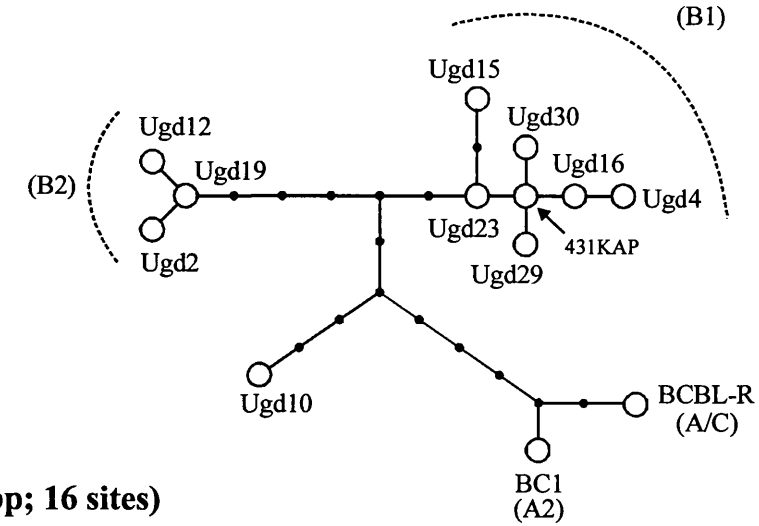
Fig. 6.6. Modified networks for (A) ORF26, (B) T0.7/K12 and (C) ORF75.

The corresponding networks in Fig. 6.3 were modified based on the data shown in Fig. 6.4 to show the relationships between the Ugandan sequences (including Ugd30) and BCBL-R and 431KAP. The length of the alignments and the number of substitution sites used to construct the networks are indicated. The groups defined previously are indicated in parentheses. Rory Bowden drew the networks in Coreldraw. The network in (A) was drawn at twice the scale as those in (B) and (C).

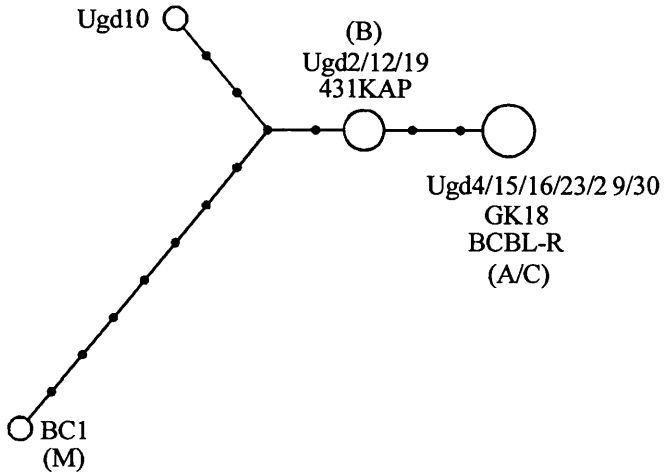
A ORF26 (298 bp; 6 sites)



B T0.7/K12 (551 bp; 27 sites)



C ORF75 (578 bp; 16 sites)



have genomes that may reflect a history of recombination between sublineages of the B subtype. Ugd2, Ugd19 and Ugd29 appear to have subtype B characteristics throughout their genomes. The genome of BCBL-R was previously assumed to be non recombinant (Poole et al., 1999), but the analysis presented here indicates that this may not be the case. In agreement with the conclusion of Poole et al. (1999), the network analysis identified recombinants among the Ugandan strains and also showed that all the Ugandan strains have some features of subtype B genomes.

Importantly, the study indicates that network analysis (and phylogenetic analysis where it can be performed reliably e.g with the K1 gene) is a useful tool for reliable and objective definition of parental genotypes and recombinants.

CHAPTER 7
GENERAL DISCUSSION

This chapter gives an overall discussion of the results and their implications in the evolution of HHV-8. The study has made a contribution towards elucidating some of the virological factors involved in KS epidemiology in Uganda. HHV-8 antibodies were detected in a high percentage of blood donors (74%) residing in Kampala. Furthermore, KS patients in Uganda are infected with HHV-8 strains belonging to three K1 subtypes (A5, B and C), and virtually all these strains are linked to the K15 P allele. A rare K1 B subtype strain containing a novel type of K15 M allele was identified in a single patient. Over half (5 of 9) of the strains in Ugandan KS patients are recombinants between subtypes.

7.1. HHV-8 SEROPREVALENCE IN BLOOD DONORS

Previous serological studies do not support a widespread distribution of HHV-8 in most normal populations. Seroprevalence rates have been reported to be high in African countries, including Uganda, The Gambia, Ivory Coast, Zambia and South Africa, compared to other areas of the world (Table 3.1 and reviewed by Chatlynne and Ablashi, 1999; Moore 2000; Schulz, 1998). In the current study, blood donors were tested for HHV-8 antibodies by combining two tests, an ELISA (against both lytic [ORF65] and latent [ORF73/LANA] antigens) and a LANA IF test. Together with WB analysis (to rule out non-specific ELISA reactions) the tests provide a method of HHV-8 antibody detection that is both highly sensitive and highly specific. This was an improvement on most previous studies, which used single tests (reviewed by Chatlynne and Ablashi, 1999; Jaffe and Pellett, 1999). A high seroprevalence rate (74%) was found in blood donors using this approach. This result provides additional evidence that HHV-8 infection is more widespread in the normal populations of Africa than elsewhere

in the world. The reasons for the geographic differences in seroprevalence are unknown (Jaffe and Pellett, 1999). Studies of HHV-8 seroprevalence in South Africa (Sitas et al., 1999; Sitas and Newton, 2001) suggest that seroprevalence rates decrease with increasing levels of education and are lower in whites than in blacks, suggesting that socioeconomic risk factors may be important determinants of HHV-8 transmission. However, a previous Ugandan study found that KS patients had more education and were more affluent than control patients (Ariyoshi et al., 1998).

HHV-8 seroprevalence rates among women and men blood donors were similar, as has been found by previous studies (Calabrò et al., 1998; Mayama et al., 1998; Sitas et al., 1999). Since KS in Africa is about 8-10 times as common among men as among women (Cook-Mozaffari et al., 1998), there must be a reason in addition to HHV-8 infection to explain this difference.

The results of this and previous studies of HHV-8 seroprevalence in Ugandan populations should be regarded as being preliminary given the relatively small sample sizes involved (Table 3.1; Chatlynne and Ablashi, 1999). More detailed, large-scale population-based studies using currently available highly sensitive and highly specific serological assays are required in order to determine absolute prevalence rates and risk factors involved in the transmission of HHV-8. Furthermore, although levels of HHV-8 seroprevalence are similarly high in Uganda, Zambia, The Gambia and Ivory Coast, endemic KS is common only in Uganda and Zambia, indicating that a co-factor(s) might be necessary for the development of KS.

7.2 HHV-8 SUBTYPES IN KS PATIENTS

7.2.1 Overall genotypes

The results of this and previous studies show that representatives of A5, B and C K1 subtypes circulate in Uganda, subtype B being predominant (Table 4.1; Fig. 4.6). The results of the current study strengthen the evidence that the K1 B subtype predominates in Africa (Cook et al., 1999; Meng et al., 1999; Zong et al., 1999; Lacoste et al., 2000a). A recent compilation of 75 K1 sequences originating from African patients or from individuals of African origin (but living elsewhere) indicated that A5 variants may be as prevalent and widespread in Africa as the B subtype (Lacoste et al., 2000a). A5 variants have been detected in samples from Zambia, Tanzania, Uganda, Cameroon, the Central African Republic and in Creoles, recent African immigrants to French Guiana (Kasolo et al., 1998; Cook et al., 1999, Meng et al., 1999; Zong et al., 1999; Lacoste et al., 2000a.). My study adds to the number of A5 variants identified from Ugandan samples. The single subtype C strain (Ugd23) identified in this study together with its counterpart (K1-43/Ber) (Lacoste et al., 2000a) appear to belong only nominally to the C subtype (Fig. 4.6).

Previous reports examining variability in the K15 gene, including an extensive study of more than 60 strains, largely used PCR assays to distinguish the P and M alleles (Poole et al., 1999, Lacoste et al., 2000a,b; Meng et al., 2001). PCR screening of 30 Ugandan KS tumour samples confirmed that the majority contains the P allele, as predicted by Poole et al. (1999). The K15 M allele was identified in a single sample (Ugd10) containing a subtype B K1 gene. This is the first time a K15 M allele has been identified in Eastern Africa samples. These results contrast those in the study by Lacoste et al (2000a), which showed an equal proportion of P and M alleles in samples from Central/West Africa.

Extension of the PCR study to sequence analysis of the entire K15 gene in selected strains revealed that the M and P alleles occur in two main forms, the BC-1 (M) and Ugd10 (M') forms for the M allele, and the Ugd2/12/19 (BI or P[B]) and BCBL-R (BII or P[A/C]) forms for the P allele (see Table 6.3).

The overall predominance of the P allele in most parts of the world, should it be confirmed by further genotyping and sequence analysis studies, may imply some form of selective advantage for the P allele over the M allele. The M allele, on the other hand, appears to survive and spread better in certain populations (e.g. Central/West Africa and Taiwan) than in others (e.g. Eastern Africa) (Lacoste et al., 2000a,b; Poole et al., 1999).

There is published evidence for recombination in about 20-30% of HHV-8 strains (including five out of 12 African strains) by the criterion of lack of co-segregation at multiple genetic loci (Poole et al., 1999). In agreement, the network analysis identified at least five of nine Ugandan strains as recombinants (Table 6.3). HHV-8 recombinants are thus common in nature. However, survival of the individual products of recombination is likely to be very rare, given the relatively low number of recombination events required to explain observed genome structures (Table 6.3), the lack of detectable recombination within individual loci, and the time thought to have elapsed since the subtypes diverged (of the order of 10^5 years; Hayward, 1999). Intraspecies recombination is not unique to HHV-8 among the herpesviruses; it has also been observed in EBV, the human herpesvirus most closely related to HHV-8 (Midgley et al., 2000). Poole et al. (1999), in their study of more than 60 HHV-8 genomes, noted that for some viral strains, sequence divergence at the K15 gene between the P and M subtypes extends leftwards (at a greatly reduced degree) up to T0.7/K12. A similar observation was made in Ugd10 (Table 6.3).

7.2.2 Correlation of subtype with disease

As in most previous studies (Cook et al., 1999; Meng et al., 1999, 2001; Lacoste et al., 2000a; Zong et al., 1999), no obvious association was found between K1 subtype and the clinical/epidemiological forms of KS. Similarly, no correlation was found with ethnicity or geographical location of the patient. Interestingly, however, all three P(BI) K15 subtypes (Ugd2/12/19) identified in this study originated from endemic KS patients (Table 4.2). Previous studies have examined correlation of HHV-8 subtypes with disease on the basis of single gene segments, e.g. K1 or ORF26. However, it is likely to be more fruitful to take into account overall genotypes. It is also possible that co-factor(s) may be responsible for the different clinical and epidemiological forms of KS. These could be other infectious agents, host or environmental factors. Whereas HIV-1 appears to be an important co-factor in epidemic (AIDS-associated) KS, co-factors in endemic KS remain a mystery.

7.2.3. Classification and nomenclature of HHV-8 subtypes

This study, like most previous studies (Cook et al., 1999; Meng et al., 2001; Lacoste et al., 2000a), highlights the need to redefine and name HHV-8 subtypes objectively. K1 subtypes (except subtype E) currently have two main designations as shown in Table 1.3, and this can be confusing. Further subdivisions have also been proposed in previous studies (Table 1.3), but the criteria for these designations are not well defined and therefore not uniform. Furthermore, the resulting classifications are not always supported by phylogenetic analysis. The K1 subtype B designations used in this study (Fig. 4.6) are also provisional and subject to change with rigorous testing methods and as more strains are described. Further complications in K1 phylogenetic analysis arise from detection of strains in the A/C lineage belonging to neither subtype

(e.g. K1-43/Ber; Fig. 4.6; Lacoste et al., 2000a) or only nominally to the C subtype (e.g. Ugd23 and K1-8/Dem; Fig. 4.6; Lacoste et al., 2000a). This problem reduces the usefulness of the phylogenetic distinction between the A and C subtypes, and could be resolved either by abolishing the distinction, or by introducing additional subtypes to accommodate novel strains.

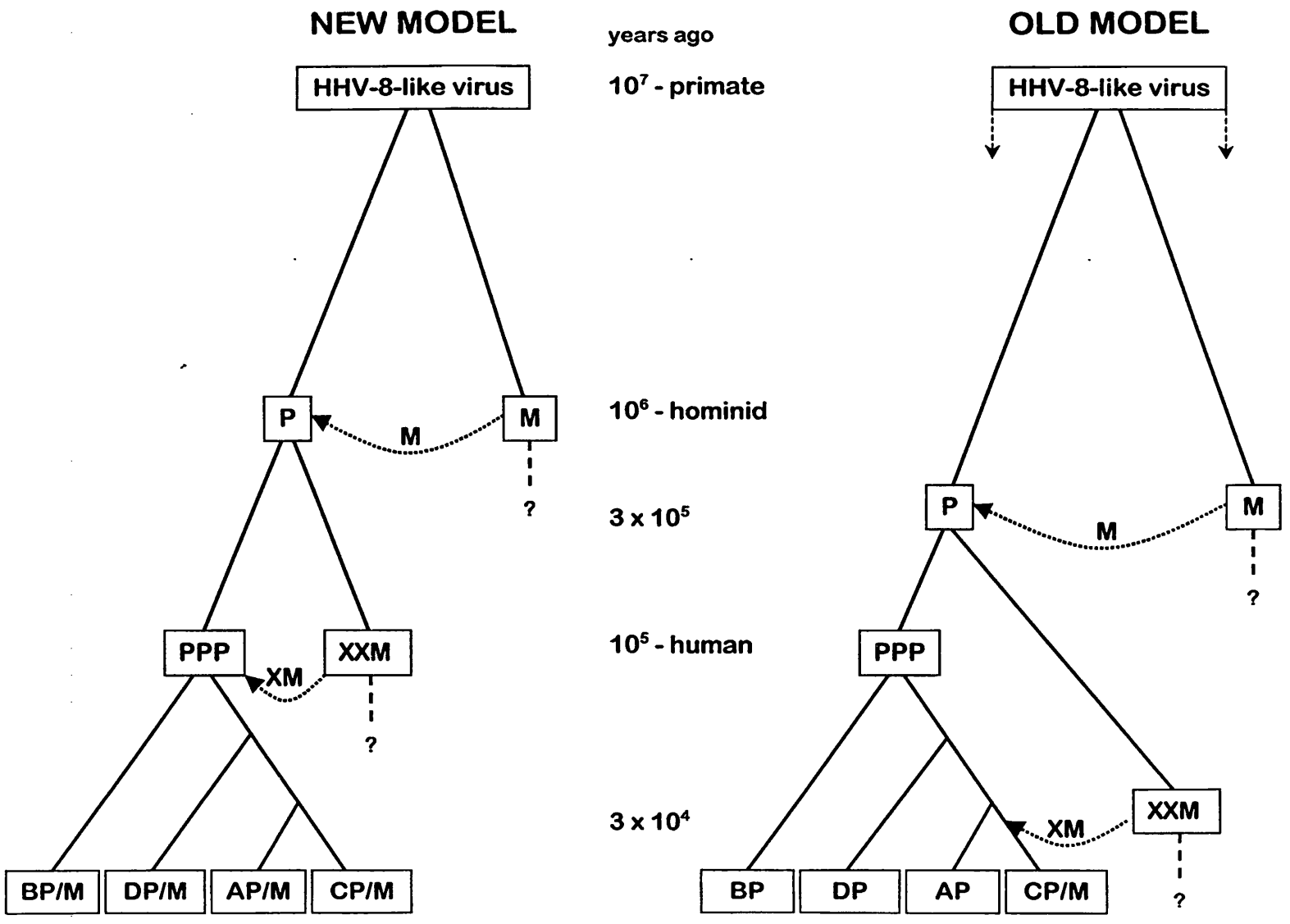
Analysis of groups at internal loci by network analysis also underlines the need to redefine parental genotypes. Previously, parental strains were determined on the basis of "weight of numbers" (Poole et al., 1999; Zong et al., 1999), but the analysis presented in chapter 6 indicates that this method might produce misleading results in some cases (e.g. BCBL-R and 431KAP which might be recombinants themselves; Fig. 6.6). Network analysis of conserved regions of the genome together with K1 phylogenetic analysis (using computer programs) of large data sets (i.e. including a large set of samples covering a large part of the genome) would help provide an objective, and therefore more reliable, definition of parental genotypes and recombinants.

7.3. EVOLUTION OF HHV-8

Whereas the hypervariability of K1 is likely to represent a rapid evolution of the modern form of HHV-8 (McGeoch and Davison, 1999b; Hayward, 1999), the very low level of nucleotide variation found within either the P or M allele suggests that the K15 divergence originated from a recombination event with a HHV-8-like virus (Hayward, 1999; Poole et al., 1999). In a model focusing on the origin of the K15 alleles (depicted in Fig. 7.1), Hayward (1999) proposed that an ancestor of HHV-8 diverged into two forms, one of which eventually led to extant P allele lineages and one of which obtained the M allele by recombination with a related Old World monkey, or perhaps great ape, virus. In order to accommodate the claim of Poole et al. (1999) that the M allele in 18 strains of

Fig. 7.1. Models for the origin of K15 divergence.

The new model is proposed in this study and the old model was proposed previously (Hayward, 1999; Poole et al., 1999). The time period for the divergence of the ancestral HHV-8 virus was not stated in the old model (reflected by arrows). Question marks indicate viruses that have not been identified or may be extinct. M and P represent M and P allele-containing strains, respectively; X, diverged P strain sequences; A, B, C and D denote K1 subtypes. An approximate time scale is shown in a non-linear fashion. I acknowledge Andrew Davison's assistance in drawing this figure.


NEW MODEL
years ago
OLD MODEL
HHV-8-like virus
10^7 - primate
HHV-8-like virus
P
M
M
?
10^6 - hominid
P
M
M
?
3×10^5
PPP
XM
XXM
?
10^5 - human
PPP
BP/M
DP/M
AP/M
CP/M
3×10^4
XM
XXM
?
BP
DP
AP
CP/M

HHV-8 exhibited virtually no sequence variation, it was proposed that the M allele and adjacent sequences re-entered extant lineages relatively recently (approximately 35000 years ago) via a single recombination event with a K1 C subtype genome. The M allele was then passed to other subtypes, with further recombination events resulting in differences in the extent of adjacent M-allele linked sequences.

The data presented here indicate that the M allele and adjacent regions have been evolving in at least two forms for a period equivalent to that during which extant subtypes have diversified. In the absence of evidence for either allele undergoing unusual evolutionary processes, this implies that introduction of the M allele into modern lineages did not occur via a single, recent event (as proposed in the model above). In principle, the M allele could have been introduced by a single, ancient recombination event into a lineage that gave rise to extant subtypes, or into modern lineages by two separate, possibly recent recombination events involving diverged donors of the M allele. Noting the high level of divergence in UPS75' between strains containing P and M alleles, the former scenario is more likely. Thus a new model can be proposed (Fig. 7.1). Two viral lineages diverged during primate evolution (of the order of 10^7 years ago), continuing to evolve with their host species and eventually producing distinct viruses containing the M or P alleles. At the stage when the host was a hominid (10^6 years ago), the M allele was transferred to a P allele-containing genome. The M allele and adjacent UPS75' region (now diverged from its equivalent in P-allele containing genomes) were then transferred into a P allele-containing lineage in an ancestral human (10^5 years ago). Further divergence, recombination and extinctions as envisaged by Hayward (1999), coupled with the vicissitudes of human survival and migration, resulted in extant HHV-8 genotypes. It is conceivable that the original M allele-containing virus corresponds to the RRV lineage present in Old

World monkeys and possibly in the great apes, but that this lineage has become extinct in humans, with only the K15 allele surviving in HHV-8 recombinants.

The identification of diverged versions of the M allele might be taken as compromising the assumption that the P allele is indigenous to HHV-8 and that the M allele was captured. Identification of apparently non-recombinant Ugandan strains that contain the P allele (Ugd2, Ugd19 and Ugd29) supports the view that the P allele was indeed indigenous, but it must be recognised that this argument depends upon correct definition of parental genotypes. Further insights into the fascinating but complex aspects of HHV-8 evolution are expected to emerge from continuing examination of extant HHV-8 lineages and related primate viruses.

7.4. CONCLUSION

Analysis of samples from Uganda for HHV-8 antibodies or HHV-8 strains has shed light on the virological factors involved in KS epidemiology in this country, as well as provided further glimpses into the evolution of HHV-8. The study has strengthened the notion that HHV-8 infection is widespread in Uganda, determined the characteristics of strains associated with KS disease, showed that the M allele (though very rare) exists in Eastern Africa, and that the P and M alleles have both been evolving in at least two forms for equivalent lengths of time. Areas for future research and refinements in currently available data have also been highlighted.

REFERENCES

Alagiozoglou L, Sitas F, Morris L. 2000. Phylogenetic analysis of human herpesvirus-8 in South Africa and identification of a novel subgroup. Journal of General Virology 81, 2029-2038.

Albini A, Soldi R, Giunciuglio D, Giraudo E, Benelli R, Primo L, Noonan D, Salio M, Camussi G, Rockl W, Bussolino F. 1996. The angiogenesis induced by HIV-1 Tat protein is mediated by the Flk-1/KDR receptor on vascular endothelial cells. Nature Medicine 2, 1371-1375.

Albrecht J-C, Nicholas J, Biller D, Cameron KR, Biesinger B, Newman C, Wittmann S, Craxton MA, Coleman H, Fleckenstein B, Honess RW. 1992. Primary structure of the Herpesvirus saimiri genome. Journal of Virology 66, 5047-5058.

Alexander L, Denekamp L, Knapp A, Auerbach MR, Damania B, Desrosiers RC. 2000. The primary sequence of rhesus monkey rhadinovirus isolate 26-95: sequence similarities to Kaposi's sarcoma-associated herpesvirus and rhesus monkey rhadinovirus isolate 17577. Journal of Virology 74, 3388-3398.

Ambroziak JA, Blackbourn DJ, Herndier BG, Glogau RG, Gullett JH, McDonald AR, Lennette ET, Levy JA. 1995. Herpes-like sequences in HIV-infected and uninfected Kaposi's sarcoma patients. Science 268, 582-583.

André S, Schatz O, Bogner JR, Zichhardt H, Stöffler-Meilicke M, Jahn H-U, Ullrich R, Sonntag A-K, Kehm R, Haas J. 1997. Detection of antibodies against viral capsid proteins of human herpesvirus 8 in AIDS-associated Kaposi's sarcoma. Journal of Molecular Medicine 75, 145-152.

Angeloni A, Heston L, Uccini S, Sirianni MC, Cottoni F, Masala MV, Cerimele D, Lin SF, Sun R, Rigsby M, Faggioni A, Miller G. 1998. High prevalence of antibodies to human herpesvirus 8 in relatives of patients with classic Kaposi's sarcoma from Sardinia. Journal of Infectious Diseases 177, 1715-1718.

Ansari MQ, Dawson DB, Nador R, Rutherford C, Schneider NR, Latimer MJ, Picker L, Knowles DM, McKenna RW. 1996. Primary body cavity-based AIDS-related lymphomas. American Journal of Clinical Pathology 105, 221-229.

Ariyoshi K, Van der Loeff MS, Corrah T, Cham F, Cook PM, Whitby D, Weiss RA, Schulz TF, Whittle H. 1998. Kaposi's sarcoma and human herpesvirus 8 (HHV8) in HIV-1 and HIV-2 infection in The Gambia. Journal of Human Virology 1, 193-199.

Arvanitakis L, Geras-Raaka E, Varma A, Gershengorn MC, Cesarman E. 1997. Human herpesvirus KSHV encodes a constitutively active G-protein-coupled receptor linked to cell proliferation. Nature 385, 347-350.

Arvanitakis L, Mesri EA, Nador R, Said JW, Asch AS, Knowles DM, Cesarman E. 1996. Establishment and characterisation of a primary effusion (body cavity-based) lymphoma cell line (BC-3) harbouring Kaposi's sarcoma-associated herpesvirus (KSHV/HHV-8) in the absence of Epstein-Barr virus. Blood 88, 2648-2654.

Athale UH, Patil PS, Chintu C, Elem B. 1995. Influence of HIV epidemic on the incidence of Kaposi's sarcoma in Zambian children. Journal of Acquired Immune Deficiency Sundromes and Human Retrovirology 8, 96-100.

Bais C, Santomasso B, Coso O, Arvanitakis L, Raaka EG, Gutkind JS, Asch AS, Cesarman E, Gershengorn MC, Mesri EA. 1998. G-protein-coupled receptor of Kaposi's sarcoma-associated herpesvirus is a viral oncogene and angiogenesis activator. Nature 391, 86-89.

Ballestas ME, Chatis PA, Kaye KM. 1999. Efficient persistence of extrachromosomal KSHV DNA mediated by latency-associated nuclear antigen. Science 284, 641-644.

Bandelt HJ, Forster P, Sykes BC, Richards MB. 1995. Mitochondrial portraits of human populations using median networks. Genetics. 141, 743-753.

Barozzi P, Luppi M, Masini L, Marasca R, Savarino M, Morselli M, Ferrari MG, Bevini M, Bonacorsi G, Torelli G. 1996. Lymphotropic herpesvirus (EBV, HHV-6, HHV-8) DNA sequences in HIV negative Castleman's disease. Journal of Clinical Pathology 49, M232-M235.

Beckstead J, Wood G, Fletcher V. 1985. Evidence of the origin of Kaposi's sarcoma from lymphatic endothelium. American Journal of Pathology 119, 294-300.

Benton MJ. 1990. Vertebrate Palaeontology. HarperCollins Academic, London.
Beral V, Peterman TA, Berkelman RL, Jaffe HW. 1990. Kaposi's sarcoma among persons with AIDS: a sexually transmitted infection? Lancet 335, 123-128.

Beral V. 1991. Epidemiology of Kaposi's sarcoma. In: Cancer, HIV and AIDS. Cancer Surveys, vol.10 (ed. Beral V, Jaffe HW, Weiss RA), pp 5-22. Cold Spring Harbor, New York.

Beral V, Peterman TA, Berkeman RL, Jaffe HW. 1990. Kaposi's sarcoma among persons with AIDS: a sexually transmitted infection? Lancet 335, 123-128.

Besnard V, Euvrad S, Kanitakis J, Mion F, Boillot O, Frances M, Faure M, Claudy A. 1996. Kaposi's sarcoma after liver transplantation. Dermatology, 193, 100-104.

Biggar RJ, Whitby D, Marshall V, Linhares AC, Black F. 2000. Human herpesvirus 8 in Brazilian Amerindians: a hyperendemic population with a new subtype. Journal of Infectious Diseases 181, 1562-1568.

Bigoni B, Dolcetti R, de Lellis L, Carbone A, Boiocchi M, Cassai E, DiLuca D. 1996. Human herpesvirus 8 is present in the lymphoid system of healthy persons

and can reactivate in the course of AIDS. Journal of Infectious diseases 173, 542-549.

Birkenbach M, Josefsen K, Yalamanchili R, Lenoir G, Kieff E. 1993. Epstein-Barr virus-induced genes: first lymphocyte-specific G protein-coupled peptide receptors. Journal of Virology 67, 2209-2220.

Blackbourn DJ, Ambroziak J, Lennette E, Adams M, Ramachandran B, Levy JA. 1997. Infectious human herpesvirus 8 in a healthy North American blood donor. Lancet 349, 609-611.

Blackbourn DJ, Levy JA. 1997. Human herpesvirus 8 in semen and prostate. AIDS 11, 249-250.

Blauvelt A, Sei S, Cook PM, Sculz TF, Jeang KT. 1997. Human herpesvirus 8 infection occurs following adolescence in the United States. Journal of Infectious Diseases 176, 771-774.

Boldough I, Szaniszlo P, Bresnahan WA, Flaitz CM, Nichols MC, Albrecht T. 1996. Kaposi's sarcoma herpesvirus-like DNA sequences in the saliva of individuals infected with human immunodeficiency virus. Clinical Infectious Diseases 23, 406-407.

Boralevi F, Masquelier B, Denayrolles M, Dupon M, Pellegrin JL, Ragnaud JM, Fleury HJ. 1998. Study of human herpesvirus 8 (HHV-8) variants from Kaposi's sarcoma in France: Is HHV-8 subtype A responsible for more aggressive tumors? Journal of Infectious Diseases 178, 1546-1547.

Boshoff C, Gao S-J, Healy LE, Matthews S, Thomas AJ, Cignet L, Warnke RA, Strauchen JA, Matutes E, Kamel OW, Moore PS, Weiss RA, Chang Y. 1998. Establishing a KSHV positive cell line (BCP-1) from peripheral blood and characterizing its growth in Nod/SCID mice. Blood 91, 1671-1679.

Boshoff C, Schulz TF, Kennedy MM, Graham AK, Fisher C, Thomas A, McGee JO, Weiss RA, O'Leary JJ. 1995a. Kaposi's sarcoma-associated herpesvirus infects endothelial and spindle cells. Nature Medicine. 1, 1274-1278.

Boshoff C, Talbot S, Kennedy M, O'Leary J, Schulz TF, Chang Y. 1996. HHV8 and skin cancers in immunosuppressed patients. Lancet 347, 338-339.

Boshoff C, Whitby D, Hatzioannou T, Fisher C, van der Walt J, Hatzakis A, Weiss RA, Schulz TF. 1995b. Kaposi's sarcoma-associated herpesvirus in HIV-negative Kaposi's sarcoma. Lancet 345, 1043-1044.

Bouquety JC, Siopathis MR, Ravisse, PR, Lagarde N, Georges-Courbot MC, Georges AJ. 1989. Lympho-cutaneous Kaposi's sarcoma in an African paediatric AIDS case. American Journal of Tropical Medicine and Hygiene 40, 323-325.

Brambilla L, Boneschi V, Berti E, Corbellino M, Parravicini C. 1996. HHV8 cell-associated viraemia and clinical presentation of Mediterranean Kaposi's sarcoma. Lancet 347, 1338.

Buonaguro FM, Tornesello ML, Beth-Giraldo E, Hatzakis A, Mueller N, Downing R, Biryahwaho B, Sempala SD, Giraldo G. 1996. Herpesvirus-like DNA sequences detected in endemic, classic, iatrogenic and epidemic Kaposi's sarcoma (KS) biopsies. International Journal of Cancer 65, 25-28.

Calabrò ML, Sheldon J, Simson GR, Fiore JR, Gomes E, Angarano G, Angarano G, Chieco-Bianchi L, Schulz TF. 1998. Seroprevalence of Kaposi's sarcoma-associated herepsevirus/Human herpesvirus 8 in several regions of Italy. Journal of Human Virology 1, 207-213.

Cannon JS, Ciufo D, Hawkins AL, Griffin CA, Borowitz MJ, Hayward GS, Ambinder RF. 2000. A new primary effusion lymphoma-derived cell line yields a highly infectious Kaposi's sarcoma herpesvirus-containing supernatant. Journal of Virology 74, 10187-10193.

Carbone A, Gloghini A, Vaccher E, Zagonel V, Pastore C, dalla Palma P, Branz F, Saglio G, Volpe R, Tirelli UG, Gaidano G. 1996. Kaposi's sarcoma-associated herpesvirus DNA sequences in AIDS-related and AIDS-unrelated lymphomatous effusions. British Journal of Haematology 94, 533-543.

Cavalli-Sforza LL, Cavalli-Sforza F. 1995. The great human diasporas: the history of diversity and evolution, Chapters 2 and 3. Addison-Wesley Publishing Company.

Cerimele F, Curreli F, Ely S, Friedman-Kien AE, Cesarman E, Flore O. 2001. Kaposi's sarcoma-associated herpesvirus can productively infect primary human keratinocytes and alter their growth properties. Journal of Virology 75, 2435-2443.

Cesarman E, Chang Y, Moore PS, Said JW, Knowles DM. 1995a. Kaposi's sarcoma-associated herpesvirus DNA sequences in AIDS-related body cavity-based lymphomas. New England Journal of Medicine 332, 1186-1191.

Cesarman E, Moore PS, Rao PH, Inghirami G, Knowles DM, Chang Y. 1995b. In vitro establishment and characterisation of two acquired immunodeficiency syndrome-related lymphoma cell lines (BC-1 and BC-2) containing Kaposi's sarcoma-associated herpesvirus-like (KSHV) DNA sequences. Blood 86, 2708-2714.

Cesarman E, Nador RG, Aozasa K, Delsol G, Said JW, Knowles DM. 1996a. Kaposi's sarcoma-associated herpesvirus in non-AIDS related lymphomas occurring in body cavities. American Journal of Pathology 149, 53-57.

Cesarman E, Nador RG, Bai F, Bohenzky RA, Russo JJ, Moore PS, Chang Y, Knowles DM. 1996b. Kaposi's sarcoma-associated herpesvirus contains G protein-coupled receptor and cyclin D homologs which are expressed in Kaposi's sarcoma and malignant lymphoma. Journal of Virology 70, 8218-8223.

Cha TA, Tom E, Kemble GW, Duke GM, Mocarski ES, Spaete RR. 1996. Human cytomegalovirus clinical isolates carry at least 19 genes not found in laboratory strains. Journal of Virology 70, 78-83.

Chandran B, Bloomer G, Chan SR, Zhu L, Goldstein E, Horrat R. 1998a. Human herpesvirus-8 ORF K8.1 gene encodes immunogenic glycoproteins generated by spliced transcripts. Virology 249, 140-149.

Chandran B, Smith MS, Koelle DM, Cory L, Horvat R, Goldstein E. 1998b. Reactivities of human sera with human herpesvirus-8 in infected BCBL-1 cells and identification of HHV-8-specific proteins and glycoproteins and the encoding cDNAs. Virology 243, 208-217.

Chang Y, Cesarman E, Pessin MS, Lee F, Culpepper J, Knowles DM, Moore PS. 1994. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. Science 266, 1865-1869.

Chang Y, Moore PS, Talbot SJ, Boshoff C, Zarkowska CT, Godden-Kent D, Paterson H, Weiss RA, Mittnacht S. 1996a. Cyclin encoded by KS herpesvirus. Nature 382, 410.

Chang Y, Ziegler J, Wabinga H, Katongole-Mbidde E, Boshoff C, Schulz TF, Whitby D, Maddalena D, Jaffe HW, Weiss RA, Moore PS. 1996b. Kaposi's sarcoma-associated herpesvirus DNA sequences are present in African endemic and AIDS-associated Kaposi's sarcoma. Archives of Internal Medicine 156, 202-204.

Chatlynne LG, Ablashi DV. 1999. Seroepidemiology of Kaposi's sarcoma-associated herpesvirus (KSHV). Seminars in Cancer Biology. 9, 175-185.

Chee MS, Bankier AT, Beck S, Bohni R, Brown CM, Cerny R, Horsnell T, Hutchison III CA, Kouzarides T, Martignetti JA, Preddie E, Satchwell SC, Tomlinson P, Weston KM, Barrell BG. 1990. Analysis of the protein coding

content of the sequence of human cytomegalovirus strain AD169. Current Topics in Microbiology and Immunology 154, 125-169.

Choi J-K, Lee B-S, Shim S-N, Li M, Jung JU. 2000. Identification of the novel K15 gene at the rightmost end of the Kaposi's sarcoma-associated herpesvirus genome. Journal of Virology 74, 436-446.

Chuck S, Grant RM, Katongole-Mbidde E, Conant M, Ganem D. 1996. Frequent presence of a novel herpesvirus genome in lesions of human immunodeficiency virus-negative Kaposi's sarcoma. Journal of Infectious Diseases 173, 248-251.

Comps M, Cochennec N. 1993. A herpes-like virus from the European oyster ostrae-edulis L. Journal of Invertebrate Pathology 62, 201-203.

Cook PM, Whitby D, Calabrò M-L, Luppi M, Kakoola DN, Hjalgrim H, Ariyoshi K, Ensoli B, Davison AJ, Schulz TF, the International Collaborative group. 1999. Variability and evolution of Kaposi's sarcoma-associated herpesvirus in Europe and Africa. AIDS, 13, 1165-1176.

Cook-Mozaffari P, Newton R, Beral V, Burkitt DP. 1998. The geographical distribution of Kaposi's sarcoma and of lymphomas in Africa before the AIDS epidemic. British Journal of Cancer 78, 1521-1528.

Corbellino C, Bestetti G, Galli M, Parravicini C. 1996a. Absence of HHV-8 in prostate and semen. New England Journal of Medicine 335, 1237.

Corbellino M, Poirel L, Aubin JT, Paulli M, Magrini U, Bestetti G, Galli M, Parravicini C. 1996b. The role of human herpesvirus 8 and Epstein-Barr virus in the pathogenesis of giant lymph node hyperplasia (Castleman's disease). Clinical Infectious Diseases 22, 1120-1121.

Corbellino M, Poirel L, Bestetti G, Pizzuto M, Aubin JT, Capra M, Bifulco C, Berti E, Agut H, Rizzardini G, Galli M, Parravicini C. 1996c. Restricted tissue distribution of extralesional Kaposi's sarcoma-associated herpesvirus-like DNA

sequences in AIDS patients with Kaposi's sarcoma. AIDS Research and Human Retroviruses 12, 651-657.

Coscoy l, Ganem D. 2000. Kaposi's sarcoma-associated herpesvirus encodes two proteins that block cell surface display of MHC class 1 chains by enhancing their endocytosis. Proceedings of the National Academy of Sciences, USA 97, 8051-8056.

Damania B, Li M, Choi J-K, Alexander L, Jung JU, Desrosiers RC. 1999. Identification of the R1 oncogene and its protein product from the rhadinovirus of rhesus monkeys. Journal of Virology 73, 5123-5131.

Dambaugh T, Hennessey K, Chamnankit L, Kieff E. 1984. U2 region of Epstein-Barr virus DNA may encode Epstein-Barr virus nuclear antigen 2. Proceedings of the National Academy of Sciences, USA 81, 7632-7636.

Dargan DJ, Jamieson FE, Maclean J, Dolan A, Addison C, McGeoch DJ. 1997. The published DNA sequence of human cytomegalovirus AD169 lacks 929 base pairs affecting genes UL42 and UL43. Journal of Virology 71, 9833-9836.

Davidovici B, Karakis I, Bourboulia D, Ariad S, Zong J-C, Benharroch D, Dupin N, Weiss R, Hayward G, Sarov B, Boshoff C. 2001. Seroepidemiology and molecular epidemiology of Kaposi's sarcoma-associated herpesvirus among Jewish population groups in Israel. Journal of the National Cancer Institute 93, 194-202.

Davis DA, Humphrey RW, Newcomb FM, O'Brien TR, Goeder JJ, Straus SE, Yarchoan R. 1997. Detection of serum antibodies to a Kaposi's sarcoma-associated herpesvirus-specific peptide. Journal of Infectious Diseases 175, 1071-1079

Davison AJ, Sauerbier E, Dolan A, Addison C, McKinnell RG. 1999. Genomic studies of the Lucké tumor herpesvirus (RaHV-1). Journal of Cancer Research and Clinical Oncology 125, 232-238.

Davison AJ. 1992. Channel catfish virus: a new type of herpesvirus. Virology 186, 9-14.

Davison AJ. 1998. The genome of salmonid herpesvirus 1. Journal of Virology 72, 1974-1982.

Davison AJ. 2001. Evolution of herpesviruses. Journal of Veterinary Microbiology. (Submitted).

Decker LL, Shankar P, Khan G. 1996. The Kaposi's sarcoma-associated herpesvirus (KSHV) is present as an intact latent genome in KS tissue but replicates in the peripheral blood mononuclear cells of KS patients. Journal of Experimental Medicine 184, 283-288.

Delabesse E, Oksenhendler E, Lebbe C, Verola O, Varet B, Turhan AG. 1997. Molecular analysis of clonality in Kaposi's sarcoma. Journal of Clinical Pathology 50, 664-668.

Desrosiers RC, Sasseviille VG, Czajak SC, Zhang X, Mansfield KG, Kaur A, Johnson RP, Lackner AA, Jung JU. 1997. A herpesvirus of rhesus monkeys related to the human Kaposi's sarcoma-associated herpesvirus. Journal of Virology 71, 9764-9769.

Diaz-Cano SJ, Wolfe HJ. 1997. Clonality in Kaposi's sarcoma. New England Journal of Medicine 337, 571.

Dictor M, Rambech E, Way D, Witte M, Bendsoe N. 1996. Human herpesvirus 8 (Kaposi's sarcoma-associated herpesvirus) DNA in Kaposi's sarcoma lesions, AIDS Kaposi's sarcoma cell lines, endothelial Kaposi's sarcoma simulators, and the skin of immunosuppressed patients. American Journal of Pathology 148, 2009-2016.

Djerbi M, Screpanti V, Catrina AI, Bogen B, Biberfeld P, Grandien A. 1999. The inhibitor of death receptor signaling, FLICE-inhibitory protein defines a new class of tumor progression factors. Journal of Experimental Medicine 190, 1025-1032.

Dupin N, Fisher C, Kellam P, Ariad S, Tulliez M, Franck N, Marck EV, Salmon D, Gorin I, Escande J-P, Weiss RA, Alitalo K, Boschoff C. 1999. Distribution of human herpesvirus-8 latently infected cells in Kaposi's sarcoma, multicentric Castleman's disease, and primary effusion lymphoma. Proceedings of the National Academy of Sciences, USA 96, 4546-4551.

Dupin N, Grandadam M, Calvez V, Gorin I, Aubin JT, Harvard S, Lamy F, Leibowitch M, Huraux JM, Escande JP, Agut H. 1995. Herpesvirus-like DNA sequences in patients with Mediterranean Kaposi's sarcoma. Lancet 345, 761-762.

Ensoli B, Gendelman R, Markham P, Fiorelli V, Colombini S, Raffeld M, Cafaro A, Chang H, Brady J, Gallo RC. 1994. Synergy between basic fibroblast growth factor and HIV-tat protein in induction of Kaposi's sarcoma. Nature 371, 674-680.

Ensoli B, Nakamura S, Salahuddin SZ, Bieberfeld P, Larsson L, Beaver B, Wong-Staal F, Gallo RC. 1989. AIDS-Kaposi's sarcoma-derived cells express cytokines with autocrine and paracrine growth effects. Science 243, 223-226.

Fenoglio CM, Oster MW, LoGerfo P, Reynolds T, Edelson R, Patterson JAK, Madeiros E, McDougall JK. 1982. Kaposi's sarcoma following chemotherapy for testicular cancer in a homosexual man: Demonstration of cytomegalovirus RNA in sarcoma cells. Human Pathology 13, 955-959.

Finke J, Fritzen R, Ternes P, Trivedi P, Bross KJ, Lange W, Mertelsmann R, Dolken G. 1992. Expression of bcl-2 in Burkitt's lymphoma cell lines: induction by latent Epstein-Barr virus genes. Blood 80, 459-469.

Finke J, Lange W, Mertelsmann R, Dolken G. 1994. Bcl-2 induction is part of the strategy of Epstein-Barr virus. Leukemia Lymphoma 12, 413-419.

Flore O, Rafii S, Ely S, O'Leary JJ, Hyjek EM, Cesarman E. 1998. Transformation of primary human endothelial cells by Kaposi's sarcoma-associated herpesvirus. Nature 394, 588-592.

Foreman KE, Friborg J, Kong WP, Woffendin C, Polverion PJ, Nickoloff BJ, Nabel GJ. 1997. Propagation of a human herpesvirus from AIDS-associated Kaposi's sarcoma. New England Journal of Medicine 336, 163-171.

Fotsis T, Zhang Y, Pepper MS, Adlercreutz H, Montesano R, Nawroth PP, Schweigerer L. 1994. The endogenous oestrogen metabolite 2-methoxyoestradiol inhibits angiogenesis and suppresses tumour growth. Nature, 368, 237-239.

Friborg J Jr, Kong W, Hottiger MO, Nabel GJ. 1999. p53 inhibition by the LANA protein of KSHV protects against cell death. Nature 402, 889-894.

Friborg Jr J, Kong W-P, Flowers CC, Flowers SL, Sun Y, Foreman KE, Nickoloff BJ, Nabel GJ. 1998. Distinct biology of Kaposi's sarcoma-associated herpesvirus from primary lesions and body cavity lymphomas. Journal of Virology 72, 10073-10082.

Friedman-Kien AE, Saltzman BR, Cao Y, Nestor MS, Mirabile M, Li JJ, Peterman TA. 1990. Kaposi's sarcoma in HIV-negative homosexual men. Lancet 335, 168-169.

Gaidano G, Cechova K, Chang Y, Moore PS, Knowles DM, Dalla-Favera R. 1996a. Establishment of AIDS-related lymphoma cell lines from lymphomatous effusions. Leukemia 10, 1237-1240.

Gaidano G, Pastore C, Gloghini A, Cusini M, Nomdedéu J, Volpe G, Capello D, Vaccher E, Bordes R, Tirelli U, Saglio G, Carbone A. 1996b. Distribution of

human herpesvirus-8 sequences throughout the spectrum of AIDS-related neoplasia. AIDS 10, 941-949.

Ganem D. 1996. Human herpesvirus 8 and the biology of Kaposi's sarcoma. Seminars in Virology 7, 325-332.

Gao SJ, Boshoff C, Jayachandra S, Weiss RA, Chang Y, Moore PS. 1997. KSHV ORF K9 (vIRF) is an oncogene which inhibits the interferon signaling pathway. Oncogene 15, 1979-1985.

Gao SJ, Kingsley L, Hoover DR, Spira TJ, Rinaldo CR, Saah A, Phair J, Detels R, Parry P, Chang Y, Moore PS. 1996a. Seroconversion of antibodies to Kaposi's sarcoma-associated herpesvirus-related latent nuclear antigens before the development of Kaposi's sarcoma. New England Journal of Medicine 335, 233-241.

Gao SJ, Kingsley L, Li M, Zheng W, Parravicini C, Ziegler J, Newton R, Rinaldo CR, Saah A, Phair J, Detels R, Chang Y, Moore PS. 1996b. KSHV antibodies among North Americans, Italians, and Ugandans with and without Kaposi's sarcoma. Nature Medicine 2, 925-928.

Gessain A, Brière J, Angelin-Duclos C, Valensi F, Béral HM, Davi F, Nicola M-A, Sudaka A, Fouchard N, Gabarre J, Troussard X, Dulmet E, Audouin J, Diebold J, de Thé, G. 1997. Human herpesvirus 8 (Kaposi's sarcoma virus) and malignant lymphoproliferations in France: a molecular study of 250 cases including two AIDS-associated body cavity-based lymphomas. Leukemia, 11, 266-272.

Gessain A, Sudaka A, Briere J, FouchardN, Nicola MA, Rio B, Arborio M, Troussard X, Audouin J, Diebold J, de The, G. 1996. Kaposi's sarcoma-associated herpes-like virus (human herpesvirus 8) DNA sequences in Multicentric Castleman's disease: is there any relevant association in non-human immunodeficiency virus-infected patients? Blood 87, 414-416.

Glenn M, Rainbow L, Aurade F, Davison A, Schulz TF. 1999. Identification of a spliced gene from Kaposi's sarcoma-associated herpesvirus encoding a protein with similarities to latent membrane proteins 1 and 2A of Epstein-Barr virus. Journal of Virology 73, 6953-6963.

Glesby MJ, Hoover DR, Weng S, Graham NM, Phair JP, Detels R, Ho M, Saah AJ. 1996. Use of antiherpes drugs and the risk of Kaposi's sarcoma from the multicenter AIDS cohort study. Journal of Infectious Diseases 173, 1477-1480.

Gompels UA, Kasolo FC. 1996. HHV-8 serology and Kaposi's sarcoma. Lancet 348, 1578-1588.

Greensill J, Thomas TF. 2000. Rhadinoviruses (gamma2-herpesviruses) of Old World primates: models for KSHV/HHV8-associated disease? AIDS 14 (Supplement 3), S11-19.

Guo HG, Browning P, Nicholas J, Hayward GS, Tschachler E, Jiang YW, Sadowska M, Raffeld M, Colombini S, Gallo RC, Reitz Jr MS. 1997. Characterization of a chemokine receptor-related gene in human herpesvirus 8 and its expression in Kaposi's sarcoma. Virology 228, 371-378.

Gupta P, Singh MK, Rinaldo C, Ding M, Farzadegan H, Saah A, Hoover D, Moore PS, Kingsley L. 1996. Detection of Kaposi's sarcoma herpesvirus DNA in semen of homosexual men with Kaposi's sarcoma. AIDS 10, 1596-1598.

Gyulai R, Kemény L, Ádám E, Nagy F, Dobozy A. 1996a. HHV8 DNA in angiolymphoid hyperplasia of the skin. Lancet 347, 1837.

Gyulai R, Kemény L, Kiss M, Ádám E, Nagy F, Dobozy A. 1996b. Herpesvirus-like DNA sequence in angiosarcoma in a patient without HIV infection. New England Journal of Medicine 334, 540-541.

Hall S, Padgett R. 1996. Requirement of U12 snRNA for *in vivo* splicing of a minor class of eukaryotic nuclear pre-mRNA introns. Science 271, 1716-1718.

Harrington WJ, Bagasra O, Sosa CE, Bobroski LE, Baum M, Wen-XL, Cabral L, Byrne GE, Pomerantz RJ, Wood C. 1996. Human herpesvirus type 8 DNA sequences in cell-free plasma and mononuclear cells of Kaposi's sarcoma patients. Journal of Infectious Diseases 174, 1101-1105.

Harwood AR, Osoba D, Hofstader SL. 1979. Kaposi's sarcoma in recipients of renal transplants. American Journal of Medicine 67, 759-765.

Hayward GS. 1999. KSHV strains: the origins and global spread of the virus. Seminars in Cancer Biology. 9, 187-199.

Hermine O, Michel M, Buzyn-Veil A, Gessain. 1996. Body cavity-based lymphoma in an HIV-seronegative patient without Kaposi's sarcoma-associated herpesvirus-like DNA sequences. New England Journal of Medicine 334, 272-273.

Hjalgrim H, Melbye M, Lecker S, Frisch M, Thomsen HK, Larsen SO. 1996. Epidemiology of classic Kaposi's sarcoma in Denmark between 1970 and 1992. Cancer (Philadelphia) 77, 1373-1378.

Homa FL, Brown JC. 1997. Capsid assembly and DNA packaging in herpes simplex virus. Reviews in Medical Virology 7, 107-122.

Howard MR, Whitby D, Bahadur G, Suggett F, Boshoff C, tenant FM, Schulz TF, Kirk S, Matthews S, Weller L, Tedder RS, Weiss RA. 1997. Detection of human herpesvirus 8 DNA in semen from HIV-infected individuals but not from healthy semen donors. AIDS, 11, F15-F19.

Humphrey RW, O'Brien TR, Newcomb FM, Nishihara H, Wyvill KM, Ramos GA, Saville MW, Goedert JJ, Straus SE, Yarchoan R. 1996. Kaposi's sarcoma (KS)-associated herpesvirus-like DNA sequences in peripheral blood

mononuclear cells: association with KS and persistence in patients receiving anti-herpesvirus drugs. Blood 88, 297-301.

Iochim HL. 1995. Kaposi's sarcoma and KSHV. Lancet 346, 1360.

Ishido S, Wang C, Lee B-S, Cohen GB, Jung JU. 2000. Downregulation of major histocompatibility complex class 1 molecules by Kaposi's sarcoma-associated herpesvirus K3 and K5 proteins. Journal of Virology 74, 5300-5309.

Jaffe ES. 1996. Primary body cavity-based AIDS-related lymphomas: evolution of a new disease entity. American Journal of Clinical Pathology 105, 141-143.

Jaffe HW, Pellett PE. 1999. Human herpesvirus 8 and Kaposi's sarcoma - some answers, more questions. New England Journal of Medicine. 340, 1912-1913.

Jenner RG, Albà MM, Boshoff C, Kellam P. 2001. Kaposi's sarcoma-associated herpesvirus latent and lytic gene expression as revealed by DNA arrays. Journal of Virology 75, 891-902.

Jin Y-T, Tsai ST, Yan J-J, Hsiao J-H, Lee Y-Y, Su I-J. 1996. Detection of Kaposi's sarcoma-associated herpesvirus-like DNA sequence in vascular lesions. American Journal of Clinical Pathology 105, 360-363.

Jones JL, Hanson DL, Chu SY, Ward JW, Jaffe HW. 1995a. AIDS-associated Kaposi's sarcoma. Science 267, 1078-1079.

Jones K, Rivera C, Sgadari C, Franklin J, Max EE, Bhatia K, Tosato G. 1995b. Infection of human endothelial cells with Epstein-Barr virus. Journal of Experimental Medicine 182, 1213-1221.

Jung JU, Desrosiers RC. 1991. Identification and characterization of the herpesvirus saimiri oncoprotein STP-C488. Journal of Virology 65, 6953-6960.

Jung JU, Trimble JJ, King N.W, Biesinger B, Fleckenstein BW, Desrosiers RC. 1991. Identification of transforming genes of subgroup A and C strains of herpesvirus saimiri. Proceedings of the National Academy of Sciences, USA 88, 7051-7055.

Kaaya EE, Parravicini C, Ordonez C, Gendelman R, Berti E, Gallo RC, Biberfeld P. 1995. Heterogeneity of spindle cells in Kaposi's sarcoma:comparison of cells in lesions and in culture. Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology 10, 295-305.

Kasolo FC, Monze M, Obel N, Anderson RA, French C, Gompels UA. 1998. Sequence analysis of human herpesvirus-8 strains from both African human immunodeficiency virus-negative and -positive childhood endemic Kaposi's sarcoma show a close relationship with strains identified in febrile children and high variation in the K1 glycoprotein. Journal of General Virology 79, 3055-3065.

Kasolo FC, Mpabalwani E, Gompels UA. 1997. Infection with AIDS-related herpesviruses in human immunodeficiency virus-negative infants and endemic childhood Kaposi's sarcoma in Africa. Journal of General Virology 78, 847-856.

Katano H, Iwasaki T, Baba N, Terai M, Mori S, Iwamoto A, Kurata T, Sata T. 2000a. Identification of antigenic proteins encoded by human herpesvirus 8 and seroprevalence in the general population and among patients with and without Kaposi's sarcoma. Journal of Virology 74, 3478-3485.

Katano H, Sato Y, Kurata T, Mori S, Sata T. 2000b. Expression and localization of human herpesvirus 8-encoded proteins in primary effusion lymphoma, Kaposi's sarcoma, and multicentric Castleman's disease. Virology 269, 335-344.

Kedes DH, Ganem D. 1997. Sensitivity of Kaposi's sarcoma-associated herpesvirus replication to antiviral drugs. Journal of Clinical Investigations 99, 2082-2086.

Kedes DH, Operskalski E, Busch M, Kohn R, Flood J, Ganem D. 1996. The seroprevalence of human herpesvirus 8 (HHV 8): Distribution of infection in Kaposi's sarcoma risk groups and evidence for sexual transmission. Nature Medicine 2, 918-924.

Kellam P, Boshoff C, Whitby D, Matthews S, Weiss RA, Talbot SJ. 1997. Identification of a major latent nuclear antigen (LNA-1) in the human herpesvirus-8 genome. Journal of Human Virology 1, 19-29.

Kestens L, Melbye M, Biggar RJ, Stevens WJ, Piot P, De Muynck A, Taelman H, De Feyter M, Paluku L, Gigase P. 1985. Endemic African Kaposi's sarcoma is not associated with immunodeficiency. International Journal of Cancer 36, 49-54.

Kieff E, Liebowitz D. 1990. Epstein-Barr virus and its replication. In: Virology (second edition; ed. Fields BN, Knipe DM, Chanock RM, Hirsch MS, Melnick JL, Monath TP, Roizman B), pp 1889-1921. Raven Press, New York.

Kirshner JR, Staskus K, Haase A, Lagunoff M, Ganem D. 1999. Expression of the open reading frame 74 (G-protein-coupled receptor) gene of Kaposi's sarcoma (KS)-associated herpesvirus: implications for KS pathogenesis. Journal of Virology 73, 6006-6014.

Koelle DM, Huang M-L, Chandran B, Vieiri J, Piepkorn M, Corey L. 1997. Frequent detection of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) DNA in saliva of human immunodeficiency virus-infected men: clinical and immunologic correlates. Journal of Infectious Diseases 176, 94-102.

Komanduri KV, Luce JA, McGrath MS, Herndier BG, Ng VL. 1996. The natural history and molecular heterogeneity of HIV-associated primary malignant lymphomatous effusion. Journal of Acquired Immune Deficiency Syndrome and Human Retrovirology 13, 215-226.

Krithivas A, Young DB, Liao G, Greene D, Hayward SD. 2000. Human herpesvirus 8 LANA interacts with proteins of the mSin3 corepressor complex

and negatively regulates Epstein-Barr virus gene expression in dually infected PEL cells. Journal of Virology 74, 9637-9645.

Kumar S, Hedges SB. 1998. A molecular timescale for vertebrate evolution. Nature 392, 917-920.

Lacoste V, Judde J-G, Briere J, Tulliez M, Garin B, Kassa-Kelembho E, Morvan J, Couppie P, Clyti E, Vila JF, Rio B, Delmer A, Mauclere P, Gessain A. 2000a. Molecular epidemiology of Human herpesvirus 8 in Africa: Both B and A5 K1 genotypes, as well as the M and P genotypes of K14.1/K15 loci, are frequent and widespread. Virology 278, 60-74.

Lacoste V, Kadyrova E, Chistiakova I, Gurtsevitch V, Judde J-G, Gessain A. 2000b. Molecular characterization of Kaposi's sarcoma-associated herpesvirus/human herpesvirus-8 strains from Russia. Journal of General Virology 81, 1217-1222.

Lacoste V, Mauclere P, Dubreuil G, Lewis J, Georges-Courbot M-C, Gessain A. 2000c. KSHV-like herpesviruses in chimps and gorillas. Nature, 407, 151-152.

Lagunoff M, Ganem D. 1997. The structure and coding organization of the genomic termini of Kaposi's sarcoma-associated herpesvirus. Virology 236, 147-154.

Lagunoff M, Majeti R, Weiss A, Ganem D. 1999. Deregulated signal transduction by the K1 gene product of Kaposi's sarcoma-associated herpesvirus. Proceedings of the National Academy of Sciences, USA 96, 5704-5709.

Lebbé C, Pellet C, Tatoud R, Agbalika F, Dosquet P, Desgrez JP, Morel P, Calvo F. 1997. Absence of human herpesvirus 8 sequences in prostate specimens. AIDS 11, 270.

Lee B-S, Alvarez X, Ishido S, Lackner AA, Jung JU. 2000. Inhibition of intracellular transport of B cell antigen receptor complexes by the Kaposi's

sarcoma-associated herpesvirus K1. Journal of Experimental Medicine 192, 11-21.

Lee H, Guo J, Li M, Choi J-K, DeMaria M, Rosenzweig M, Jung JU. 1998a. Identification of an immunoreceptor tyrosine-based activation motif of K1 transforming protein of Kaposi's sarcoma-associated herpesvirus. Molecular and Cellular Biology 18, 5219-5228.

Lee H, Veazey R, Williams K, Li M, Guo J, Neipel F, Fleckenstein B, Lackner A, Desrosiers RC, Jung JU. 1998b. Deregulation of cell growth by the K1 gene of Kaposi's sarcoma-associated herpesvirus. Nature Medicine 4, 435-440.

Lefrere JJ, Meyohas MC, Mariott M, Meynard JL, Thauvin M, Frottier J. 1996. Detection of human herpesvirus 8 DNA sequences before the appearance of Kaposi's sarcoma in human immunodeficiency virus (HIV)-positive subjects with a known date of HIV seroconversion. Journal of Infectious Diseases 174, 283-287.

Lennette ET, Blackbourn DJ, Levy JA. 1996. Antibodies to human herpesvirus type 8 in the general population and in Kaposi's sarcoma patients. Lancet 348, 858-861.

Levy JA, Ziegler JL (1983). Acquired immunodeficiency syndrome is an opportunistic infection and Kaposi's sarcoma results from secondary immune stimulation. Lancet 2, 78-81.

Li JJ, Huang YQ, Cockerell CJ, Friedman-Kien AE. 1996. Localization of human herpes-like virus type 8 in vascular endothelial cells and perivascular spindle-shaped cells of Kaposi's sarcoma lesions by in situ hybridization. American Journal of Pathology 148, 1741-1748.

Li M, Lee H, Guo J, Neipel F, Fleckenstein B, Ozato K, Jung JU. 1998. Kaposi's sarcoma-associated herpesvirus viral interferon regulatory factor. Journal of Virology 72, 5433-5440.

Li M, MacKey J, Czajak SC, Desrosiers RC, Lackner AA, Jung JU. 1999. Identification and characterization of Kaposi's sarcoma-associated herpesvirus K8.1 virion glycoprotein. Journal of Virology 73, 1341-1349.

Lin S-F, Sun R, Heston L, Gradoville L, Shedd D, Haglund K, Rigsby M, Miller G. 1997. Identification, expression, and immunogenicity of Kaposi's sarcoma-associated herpesvirus-encoded small viral capsid antigen. Journal of Virology 71, 3069-3076.

Liu S, Pavlova IV, Virgin HWT, Speck SH. 2000. Characterization of gammaherpesvirus 68 gene 50 transcription. Journal of Virology 74, 2029-2037.

Lubyova B, Pitha PM. 2000. Characterization of a novel human herpesvirus 8-encoded protein, vIRF-3, that shows homology to viral and cellular interferon regulatory factors. Journal of Virology 74, 8194-8201.

Lukac DM, Kirshner JR, Ganem D. 1999. Transcription activation by the product of open reading frame 50 of Kaposi's sarcoma-associated herpesvirus is required for lytic viral reactivation in B cells. Journal of Virology 73, 9348-9361.

Lukac DM, Renne R; Kirshner JR, Ganem D. 1998. Reactivation of Kaposi's sarcoma-associated herpesvirus infection from latency by expression of the ORF 50 transactivator, a homolog of the EBV R protein. Virology 252, 304-312.

Luppi M, Barozzi P, Maiorana A, Artusi T, Trovata R, Marasca R, Savarino M, Ceccherini-Nelli L, Torelli G. 1996a. Human herpesvirus-8 DNA sequences in human immunodeficiency virus-negative angioimmunoblastic lymphadenopathy and benign lymphadenopathy with giant germinal center hyperplasia and increased vascularity. Blood 87, 3903-3909.

Luppi M, Barozzi P, Maiorana A, Collina G, Grazia-Ferrari M, Matrasca R, Morselli M, Rossi E, Ceccherina-Nelli L, Torelli G. 1996b. Frequency and distribution of herpesvirus-like DNA sequences (KSHV) in different stages of

classic Kaposi's sarcoma and in normal tissues from an Italian population. International Journal of Cancer 66, 427-431.

Luppi M, Barozzi P, Marasca R, Ferrari MG, Torelli G. 1997. Human herpesvirus 8 strain variability in clinical conditions other than Kaposi's sarcoma (letter). Journal of Virology 71, 8082-8083.

Luppi M, Barozzi P, Schulz TF, Setti G, Staskus K, Trovata R, Narni F, Donelli A, Maiorana A, Marasca R, Sandrini S, Torelli G. 2000. Bone marrow failure associated with human herpesvirus 8 infection after transplantation. New England Journal of Medicine 343, 1378-1385.

Luppi M, Whitby D, Barozzi P, Boshoff C, Weiss R, Cucci F, Torelli G. 1998. Response to Manns et al. (1998). Journal of the National Cancer Institute 90, 1103-1104.

MacKenzie J, Sheldon J, Morgan G, Cook G, Schulz TF, Jarret RF. 1997. HHV-8 and multiple myeloma in the UK. Lancet 350, 1144-1145..

Manet E, Gruffat H, Trescol BMC, Moreno N, Chambard P, Giot JF, Sergeant A. 1989. Epstein-Barr virus bicistronic mRNAs generated by facultative splicing code for two transcriptional trans-activators. EMBO Journal 8, 1819-1826.

Manns A, Strickler HD, Hanchard B, Manassaram DM, Waters D, Ablashi DV. 1998. Age- and sex-specific seroprevalence of human herpesvirus 8 in Jamaica. Journal of the National Cancer Institute 90, 1102-1103.

Marcelin A-G, Dupin N, Bouscary D, Bossi P, Cacoub P, Ravaud P, Calvez V. 1997. HHV-8 and multiple myeloma in France. Lancet 350, 1144.

Martin JN, Ganem DE, Osmond DH, Page-Shafer KA, Macrae D, Kedes DH. 1998. Sexual transmission and the natural history of human herpesvirus 8 infection. New England Journal of Medicine 338, 948-954.

Masood R, Zheng T, Tulpule A, Arora N, Chatlynne L, Handy M, Whitman JJ, Kaplan M, Dosik M, Ablashi D, Gill P. 1997. Kaposi's sarcoma-associated herpesvirus and multiple myeloma. Science 278, 1970-1971.

Mayama S, Cuevas LE, Sheldon J, Omar OH, Smith DH, Okong P, Silvel B, Hart CA, Schulz TF. 1998. Prevalence and transmission of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) in Ugandan children and adolescents. International Journal of Cancer 77, 817-820.

McDonagh DP, Liu J, Gaffey MJ, Layfield LJ, Azumi N, Traweek ST. 1996. Detecion of Kaposi's sarcoma-associated herpesvirus-like DNA sequences in angiosarcoma. American Journal of Pathology 149, 1363-1368.

McGeoch DJ, Cook S, Dolan A, Jamieston FE, Telford EA. 1995. Molecular phylogeny and evolutionary timescale for the family of mammalian-herpesviruses. Journal of Molecular Biology 247, 443-458.

McGeoch DJ, Cook S. 1994. Molecular phylogeny of the Alphaherpesvirus subfamily and a proposed evolutionary timescale. Journal of Molecular Biology 238, 9-22.

McGeoch DJ, Davison AJ. 1999a. The molecular evolutionary history of the herpesviruses. In: Origin and evolution of viruses (ed. Domingo E, Webster R, Holland J), pp. 441-465. Academic Press, London, England.

McGeoch DJ, Davison AJ. 1999b. The descent of human herpesvirus 8. Seminars in Cancer Biology 9, 201-209.

McGeoch DJ, Dolan A, Ralph AC. 2000. Toward a comprehensive phylogeny for mammalian and avian herpesviruses. Journal of Virology 74, 10401-10406.

McGeoch DJ. 2001. Molecular evolution of the γ-Herpesvirinae. Philosphical Transactions of the Royal Society of London 356, 421-435.

Megaw AM, Rapaport D, Avidor B, Frenkel N, Davison AJ. 1998. The DNA sequence of the RK strain of human herpesvirus 7. Virology 244, 119-132.

Melbye M. Cook PM, Hjalgrim H, Begtrup K, Simpson GR, Biggar RJ, Ebbesen P, Schulz TF. 1998. Risk factors for Kaposi's sarcoma-associated herpesvirus (KSHV/HHV 8) seropositivity in a cohort of homosexual men. International Journal of Cancer 77, 543-548.

Meng Y-X, Sata T, Stamey FR, Voevodin A, Katano H, Koizumi H, Deleon M, De Cristofano MA, Galimberti R, Pellett PE. 2001. Molecular characterization of human herpesvirus 8 from Japan, Argentina and Kuwait. Journal of General Virology 82, 499-506.

Meng Y-X, Spira TJ, Bhat GJ, Birch CJ, Druce JD, Edlin BR, Edwards R, Gunthel C, Newton R, Stamey FR, Wood C, Pellett PE. 1999. Individuals from North America, Australasia, and Africa are infected with four different genotypes of human herpesvirus 8. Virology. 261, 106-119.

Mesri EA, Cesarman E, Arvanitakis L, Raffi S, Moore MAS, Posnett DN, Knowles DM, Asch AS. 1996. Human herpesvirus-8/Kaposi's sarcoma-associated herpesvirus is a new transmissible virus that infects B cells. Journal of Experimental Medicine 183, 2385-2390.

Midgley RS, Blake NW, Yao QY, Croom-Carter D, Cheung ST, Leung SF, Chan ATC, Johnson PJ, Huang D, Rickinson AB, Lee SP. 2000. Novel intertypic recombinants of Epstein-Barr virus in the Chinese population. Journal of Virology 74, 1544-1548.

Miles S, Rezai A, Salazar-Gonzalez JF, Meyden MV, Stevens RH, Logan DM, Mitsuyasu RT, Taga T, Hirano T, Kishimoto T, and Martinez-Maza O. 1990. AIDS-Kaposi's sarcoma-derived cells produce and respond to interleukin 6. Proceedings of the National Academy of Sciences, USA 87, 4068-4072.

Miller G, Heston L, Grogan E, Gradoville L, Rigsby M, Sun R, Shedd D, Kushnaryov VM, Grossberg S, Chang Y. 1997. Selective switch between latency and lytic replication of Kaposi's sarcoma herpesvirus and Epstein-Barr virus in dually infected body cavity lymphoma cells. Journal of Virology 71, 314-324.

Miller G, Rigsby MO, Heston L, Grogan E, Sun R, Metroka C, Levy JA, Gao S-J, Chang Y, Moore PS. 1996. Antibodies to butyrate-inducible antigens of Kaposi's sarcoma-associated herpesvirus in patients with HIV-1 infection. New England Journal of Medicine 334, 1292-1297.

Minson AC, Davison AJ, Desrosiers RC, Fleckenstein B, McGeoch DJ, Pellett PE, Roizman B, Studdert DMJ. 2000. Herpesviridae. In: Virus taxonomy (ed. Van Regenmortel MHV, Fauquet CM, Bishop DHL, Carstens EB, Estes MK, Lemon SM, Maniloff J, Mayo MA, McGeoch DJ, Pringle CR, Wickner RB), pp. 203-225. Academic Press, New York, N.Y.

Mocroft A, Youle M, Gazzard B, Morcinek J, Halai R, Phillips AN for the Royal Free/Chelsea and Westminister Hospitals Collaborative group. 1996. Anti-herpesvirus treatment and risk of Kaposi's sarcoma in HIV infection. AIDS 10, 1101-1105.

Monini P, de-Lellis L, Cassai E. 1996a. Absence of HHV-8 sequences in prostate and semen. New England Journal of Medicine 335, 1238-1239.

Monini P, de-Lellis L, Fabris M, Rigolin F, Cassai E. 1996b. Kaposi's sarcoma-associated herpesvirus DNA sequences in prostate tissue and human semen. New England Journal of Medicine 334, 1168-1172.

Moore P. 2000. The emergence of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8). New England Journal of Medicine 343, 1411-1413.

Moore PS, Boshoff C, Weiss RA, Chang Y. 1996a. Molecular mimicry of human cytokine and cytokine response pathway genes by KSHV. Science 274, 1739-1744.

Moore PS, Chang Y. 1995. Detection of herpesvirus-like DNA sequences in Kaposi's sarcoma lesions from persons with and without HIV infection. New England Journal of Medicine 332, 1181-1185.

Moore PS, Gao DJ, Dominguez G, Cesarman E, Lungu O, Knowles DM, Garber R, Pellett P, McGeoch DJ, Chang Y. 1996b. Primary characterization of a herpesvirus agent associated with Kaposi's sarcoma. Journal of Virology 70, 549-558.

Moore PS, Kingsley LA, Holmberg SD, Spira T, Gupta P, Hoover DR, Parry JP, Conley L, Jaffe H.W, Chang Y. 1996c. KSHV infection prior to onset of Kaposi's sarcoma. AIDS 10, 175-180.

Morfeld L, Torssander J. 1994. Long-term remission of Kaposi's sarcoma following foscarnet treatment in HIV-infected persons. Scandinavian Journal of Infectious Diseases 26, 753-754.

Moses AV, Fish KN, Ruhl R, Smith PP, Strussenberg JG, Zhu L, Chadran B, Nelson JA. 1999. Long-term infection and transformation of dermal microvascular endothelial cells by human herpesvirus 8. Journal of Virology 73, 6892-6902.

Muralidhar S, Pumfery AM, Hassani M, Sadaie MR, Kishishita M, Brady JN, Doniger JJ, Medvceczky P, Rosenthal LJ. 1998. Identification of Kaposin (open reading frame K12) as a human herpesvirus 8 (Kaposi's sarcoma-associated herpesvirus) transforming gene. Journal of Virology 72, 4980-4988.

Nador RG, Cesarman E, Chadburn A, Dawson DB, Ansari MQ, Sald J, Knowles DM. 1996. Primary effusion lymphoma: a distinct clinicopathologic entity associated with the Kaposi's sarcoma-associated herpesvirus. Blood 88, 645-656.

Nador RG, Cesarman E, Knowles DM, Said JW. 1995. Herpes-like DNA sequences in a body-cavity-based lymphoma in an HIV-negative pateint. New England Journal of Medicine 33, 943.

Neipel F, Albrecht J-C, Ensser A, Huang Y-Q, Li JJ, Friedman-Kien AE, Fleckenstein B. 1997a. Primary structure of the Kaposi's sarcoma-associated human herpesvirus 8 genome. Genbank accession no. U93872.

Neipel F, Albrecht J-C, Ensser A, Huang Y-Q, Li JJ, Friedman-Kien AE, Fleckenstein B. 1997b. Human herpesvirus 8 encodes a homolog of interleukin 6. Journal of Virology 71, 839-842.

Neipel F, Albrecht J-C, Fleckenstein B. 1997c. Cell-homologous genes in the Kaposi's sarcoma-associated rhadinovirus human herpesvirus 8: determinants of its pathogenicity? Journal of Virology 71, 4187-4192.

Nicholas J, Ruvulo V, Zong J, Ciufo D, Guo H-G, Reitz MS, Hayward GS. 1997a. A single 13-kilobase divergent locus in the Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) genome contains nine open reading frames that are homologous to or related to cellular proteins. Journal of Virology 71, 1963-1974.

Nicholas J, Ruvulo VR, Burns WH, Sandford G, Wan XY, Ciufo D, Hendrickson SB, Guo HG, Hayward GS, Reitz MS. 1997b. Kaposi's sarcoma-associated human herpesvirus-8 encodes homologues of macrophage inflammatory protein-1 and interleukin-6. Nature Medicine 3, 287-292.

Nicholas J, Zong J-C, Alcendor DJ, Ciufo DM, Poole LJ, Sarisky RT, Chiou C-J, Zhang X, Wan X, Guo H-G, Reitz MS, Hayward GS. 1998. Novel organizational features, captured cellular genes, and strain variability within the genome of KSHV/HHV8. Monographs of the National Cancer of Institute 23, 79-88.

Noel C, Hermans P, André J, Fayt I, Simonart T, Verhest A, Haot J, Burny A. 1996. Herpesvirus-like DNA sequences and Kaposi's sarcoma. Cancer 77, 2132-2136.

Oettle AG. 1962. Geographic and racial differences in the frequency of Kaposi's sarcoma as evidence of environmental or genetic causes. Acta Unio Internationale Cancrum, 18, 330-363.

Olsen SJ, Tarte K, Sherman W, Hale EE, Weiss MT, Orazi A, Klein B, Chang Y. 1998. Evidence against KSHV infection in the pathogenesis of multiple myeloma. Virus Research 2, 197-202.

Orenstein JM, Alkan S, Blauvelt A, Jeang K-T, Weinstein MD, Ganem D, Herndier B. 1997. Visualization of human herpesvirus type 8 in Kaposi's sarcoma by light and transmission electron microscopy. AIDS 11, 735-745.

Otsuki T, Kumar S, Ensoli B, Kingma DW, Yano T, Stetler-Stevenson M, Jaffe ES, Raffeld M. 1996. Detection of HHV-8/KSHV DNA sequences in AIDS-associated extranodal lymphoid malignancies. Leukemia 10, 1358-1362.

Palmero I, Holder A, Sinclair AJ, Dickson C, Peters G. 1993. Cyclin D1 and D2 are differentially expressed in human B-lymphoid cell lines. Oncogene 8, 1049-1054.

Parker BD, Bankier A, Satchwell S, Barrell B, Farrell PJ. 1990. Sequence and transcription of Raji Epstein-Barr virus DNA spanning the B95-8 deletion. Virology 179, 339-346.

Parravicini C, Chandran B, Corbellino M, Berti E, Paulli M, Moore PS, Chang Y. 2000. Differential viral protein expression in Kaposi's sarcoma-associated herpesvirus-infected diseases: Kaposi's sarcoma, primary effusion lymphoma, and multicentric Castleman's disease. American Journal of Pathology 156, 743-749.

Parravicini C, Lauri E, Baldini L, Neri A, Poli F, Sirchia G, Moroni M, Galli M, Corbellino M. 1997a. Kaposi's sarcoma-associated herpesvirus and multiple myeloma. Science 278, 1969-1970.

Parravicini C, Olsen SJ, Capra M, Poli F, Sirchia G, Gao SJ, Berti E, Nocera PS, Corbellino M. 1997b. Risk of Kaposi's sarcoma-associated herpesvirus transmission from donor allografts among Italian posttransplant Kaposi's sarcoma patients. Blood 90, 2826-2829.

Pastore C, Gloghini A, Volpe G, Nomdedeu J, Leonard E, Mazza UG, Saglio G, Carbone AG, Gaidano G. 1995. Distribution of Kaposi's sarcoma herpesvirus among lymphoid malignancies in Italy and Spain. British Journal of Haematology 91, 918-920.

Pau CP, Lam LL, Spira TJ, Black JB, Steward JA, Pellett PE, Respess RA. 1998. Mapping and serodiagnostic application of a dominant epitope within the human herpesvirus 8 ORF 65-encoded protein. Journal of Clinical Microbiology 36, 1574-1577.

Pauk J, Huang M-L, Brodie SJ, Wald A, Koelle DM, Schacker T, Celum C, Selke S, Corey L. 2000. Mucosal shedding of human hepresvirus 8 in men. New England Journal of Medicine 343, 1369-1377.

Pawson R, Catovsky D, Schulz TF. 1996. Lack of evidence of HHV-8 in mature T-cell lymphoproliferative disorders. Lancet 348, 1450-1451.

Penn I. 1983. Kaposi's sarcoma in immunosuppressed patients. Journal of Clinical Laboratory and Immunology 12, 1-10.

Peterman TA, Jaffe HW, Beral V. 1993. Epidemiological clues to the etiology of Kaposi's sarcoma. AIDS 7, 605-611.

Peterson B, Frizzera G. 1993. Multicentric Castleman's disease. Seminars in Oncology 20, 636-647.

Plancoulaine S, Abel L, van Beveren M, Tregouet DA, Joubert M, Tortevoye P, de
The G, Gessain A. 2000. Human herpesvirus 8 transmission from mother to child and between siblings in an endemic population. Lancet 356, 1062-1065.

Poole LJ, Zong J-C, Ciufo DM, Alcendor DJ, Cannon JS, Ambinder R, Orenstein JM, Reitz MS, Hayward GS. 1999. Comparison of genetic variability at multiple loci across the genomes of the major subtypes of Kaposi's sarcoma-associated herpesvirus reveals evidence for recombination and for two distinct types of open reading frame K15 alleles at the right-hand end. Journal of Virology 73, 6646-6660.

Purvis SF, Katongole-Mbidde E, Johnson JL, Leonard DCB, Byabazaire N, Luckey C, Schick HE, Wallis R, Elmets CA, Giam C-Z. 1997. High incidence of Kaposi's sarcoma-associated herpesvirus and Epstein-Barr virus in tumor lesions and peripheral blood mononuclear cells from patients with Kaposi's sarcoma in Uganda. Journal of Infectious Diseases 175, 957-950.

Quackenbush SL, Work TM, Balazs GH, Casey RN, Rovnak J, Chaves A, duToit L, Baines JD, Parrish CR, Bowser PR, Casey JW. 1998. Three closely related herpesviruses are associated with fibropapillomatosis in marine turtles. Virology 246, 392-399.

Raab MS, Albrecht JC, Birkmann A, Yaguboglu S, Lang D, Fleckenstein B, Neipel F. 1998. The immunogenic glycoprotein gp35-37 of human herpesvirus 8 is encoded by open reading frame K8.1. Journal of Virology 72, 6725-6731.

Rabkin CS, Bedi G, Musaba E, Sunkutu R, Mwansa N, Sidransky D, Biggar RJ. 1995. AIDS-related Kaposi's sarcoma is a clonal neoplasm. Clinical Cancer Research 1, 257-260.

Rabkin CS, Janz S, Lash A, Coleman AE, Musaba N, Liotta L, Biggar RJ, Zhuang ZP. 1997. Monoclonal origin of multicentric Kaposi's sarcoma lesions.

New England Journal of Medicine 336, 988-993.

Rabkin CS, Schulz TF, Whitby D, Lennette ET, Magpantay LI, Chatlynne L, Biggar RJ. 1998. Interassay correlation of human herpesvirus 8 serologic tests. HHV-8 Interlaboratory Collaborative Group. Journal of Infectious Diseases 178, 304-309.

Radkov SA, Kellam P, Boshoff C. 2000. The latent nuclear antigen of Kaposi's sarcoma-associated herpesvirus targets the retinoblastoma-E2F pathway and with the oncogene *Hras* transforms primary rat cells. Nature Medicine 6, 1121-1127.

Rady P, Yen A, Rollefson JL, Orengo I, Bruce S, Hughes TK, Trying SK. 1995. Herpesvirus-like DNA sequences in non-Kaposi's sarcoma skin lesions of transplant patients. Lancet 345, 1339-1340.

Rainbow L, Platt GM, Simpson GR, Sarid R, Gao S-J, Stoiber H, Herrington CS, Moore PS, Schulz TF. 1997. The 222-234 kD nuclear protein (LNA) of Kaposi's sarcoma-associated herpesvirus (Human herpesvirus 8) is encoded by orf73 and is a component of the latency-associated nuclear antigen. Journal of Virology 71, 5915-5921.

Regamey N, Cathomas G, Schwager M, Wernli M, Harr T, Erb P. 1998. High human herpesvirus 8 seroconversion in the homosexual population in Switzerland. Journal of Clinical Microbiology 36, 1784-1786.

Renne R, Barry C, Dittmer D, Compitello N, Brown PO, Ganem D. 2001. Modulation of cellular and viral gene expression by the latency-associated nuclear antigen of Kaposi's sarcoma-associated herpesvirus. Journal of Virology 75, 458-468.

Renne R, Blackbourn D, Whitby D, Levy J, Ganem D. 1998. Limited transmission of Kaposi's sarcoma-associated herpesvirus in cultured cells. Journal of Virology 72, 5182-5188.

Renne R, Lagunoff M, Zhong W, Ganem D. 1996a. The size and conformation of Kaposi's sarcoma-associated herpesvirus (Human herpesvirus 8) DNA in infected cells and virions. Journal of Virology 70, 8151-8154.

Renne R, Zhong W, Herndier B, McGrath M, Abbey N, Kedes D, Ganem D. 1996b. Lytic growth of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) in culture. Nature Medicine 2, 342-346.

Rettig MB, Ma HJ, Vescio RA, Põld M, Schiller G, Belson D, Savage A, Nishikubo C, Wu C, Fraser J, Said JW, Berenson JR. 1997. Kaposi's sarcoma-associated herpesvirus infection of bone marrow dendritic cells from multiple myeloma patients. Science 276, 1851-1854.

Rivas C, Thlick A-E, Parravicini C, Moore PS, Chang Y. 2001. Kaposi's sarcoma-associated herpesvirus LANA2 is a B-cell-specific latent viral protein that inhibits p53. Journal of Virology 75, 429-438.

Rixon FJ, McGeoch DJ. 1984. Detailed analysis of the mRNAs mapping in the short unique region of herpes simplex virus type 1. Nucleic Acids Research 13, 953-973.

Rock DL. 1993. The molecular basis of latent infections by alphaherpesviruses. Seminars in Virology 4, 157-165.

Roizman B, Desrosiers RC, Fleckenstein B, Lopez C, Minson AC, Studdert MJ. 1995. *Herpesviridae.* In: Virus Taxonomy. Archives of Virology (Suppl. 10), 114-127.

Roizman B, Desrosiers RC, Fleckenstein B, Lopez C, Minson AC, Studdert MJ. 1992. The family Herpesviridae: An update. Archives of Virology 123, 425-449.

Roth WK, Brandstetter H, Sturzl M. 1992. Cellular and molecular features of HIV-associated Kaposi's sarcoma. AIDS 6, 895-913.

Russo JJ, Bohenzky RA, Chien M-C, Chen J, Yan M, Maddalena D, Parry JP, Peruzzi D, Edelman IS, Chang Y, Moore PS. 1996. Nucleotide sequence of Kaposi's sarcoma-associated herpesvirus (HHV8). Proceedings of the National Academy of Sciences, USA 93, 14862-14867.

Rutgers J, Wieczorek R, Bonetti F, Kaplan KL, Posnett DN, Friedman-Kien AE, Knowles DM 1986. The expression of endothelial cell surface antigens AIDS-associated Kaposi's sarcoma: evidence for a vascular endothelial cell origin. American Journal of Pathology 122, 493-499.

Said W, Chien K, Takeuchi S, Tasaka T, Asou H, Cho SK, de Vos S, Cesarman E, Knowles DM, Koeffler HP. 1996. Kaposi's sarcoma-associated herpesvirus (KSHV or HHV8) in primary effusion lymphoma cells. Blood 87, 4937-4943.

Sample J, Liebowitz D, Kieff E. 1989. Two related Epstein-Barr virus membrane proteins are encoded by separate genes. Journal of Virology 63, 2743-2753.

Sample J, Young L, Martin B, Chatman T, Kieff E, Rickinson A, Kieff E. 1990. Epstein-Barr virus types 1 and 2 differ in their EBNA-3A, EBNA-3B, EBNA-3C genes. Journal of Virology 64, 4084-4092.

Sander CA, Simon M, Puchta U, Raffeld M, Kind P. 1996. HHV-8 in lymphoproliferative lesions in skin. Lancet 348, 475-476.

Sarid R, Flore O, Bohenzky RA, Chang Y, Moore PS. 1998. Transcription mapping of the Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) genome in a body cavity-based lymphoma cell line (BC-1). Journal of Virology 72, 1005-1012.

Sarid R, Olsen SJ, Moore PS. 1999. Kaposi's sarcoma-associated herpesvirus: epidemiology, virology, and molecular biology. Advances in Virus Research 52, 139-232.

Sarid R, Sato T, Bohenzky RA, Russo JJ, Chang Y. 1997. Kaposi's sarcoma-

associated herpesvirus encodes a functional bcl-2 homologue. Nature Medicine 3, 293-298.

Schalling M, Ekman M, Kaaya EE, Linde A, Biberfeld P. 1995. A role for the new herpesvirus (KSHV) in different forms of Kaposi's sarcoma. Nature Medicine 1, 707-708.

Schulz TF. 1998. Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8). Journal of General Virology 79, 1573-1591.

Schwam DR, Luciano RL, Mahajan SS, Wong L, Wilson AC. 2000. Carboxy terminus of human herpesvirus 8 latency-associated nuclear antigen mediates dimerization, transcriptional repression, and targeting to nuclear bodies. Journal of Virology 74, 8532-8540.

Searles RP, Bergquam EP, Axthelm MK, Wong SW. 1999. Sequence and genomic analysis of a rhesus macaque rhadinovirus with similarity to Kaposi's sarcoma-associated herpesvirus/human herpesvirus 8. Journal of Virology 73, 3040-3053.

Simpson GR, Schulz TF, Whitby D, Cook PM, Boshoff C, Rainbow L, Howard MR, Gao SJ, Bohenzky RA, Simmonds P, Lee C, deRuiter A, Hatzakism A, Teddar RS, Wellar IVD, Weiss RA, Moore PS. 1996. Prevalence of Kaposi's sarcoma-associated herpesvirus infection measured by antibodies to recombinant capsid protein and latent immunofluorescence antigen. Lancet 348, 1133-1138.

Siranni MC, Vincenzi L, Topino S, Scala E, Angeloni A, Gonnella R, Uccini S, Faggioni A. 1997. Human herpesvirus 8 DNA sequences in $CD8^+$ T cells. Journal of Infectious Diseases 176, 84-93.

Sitas F, Carrara H, Beral V, Newton R, Reeves G, Bull D, Jentsch U, Pacella-Norman R, Bourboulia D, Whitby D, Boshoff C, Weiss R, Patel M, Ruff P, Bezwoda WR, Retter E, Hale M. 1999. Antibodies against human herpesvirus 8

in black South African patients with cancer. New England Journal of Medicine. 340, 1863-1871.

Sitas F, Newton R. 2001. Kaposi's sarcoma in South Africa. Jouranl of National Cancer Institute Monographs 28, 1-4.

Smith MS, Bloomer C, Horvat R, Goldstein E, Casparian JM, Chandran B. 1997. Detection of human herpesvirus 8 DNA in Kaposi's sarcoma lesions and peripheral blood of human immunodeficiency virus-postive patients and correlation with serologic measurements. Journal of Infectious Diseases 176, 84-93.

Soulier J, Grollet L, Oksenhendler E, Cacoub P, Cazals-Hatem D, Babinet P, d'Agay MF, Clauvel JP, Raphael M, Degos L, Sigaux F. 1995. Kaposi's sarcoma-associated herpesvirus-like DNA sequences in multicentric Castleman's disease. Blood 86, 1276-1280.

Staden R, Beal KF, Bonfield JK. 1998. The Staden package. Computer methods in molecular biology. In: Bioinformatics methods and protocols (ed. Misener S, Krawetz SA), vol. 132, pp 115-132. The Humana Press Inc., Totowa, NJ 07512.

Staden R. 1987. Computer handling of DNA sequencing projects. In: Nucleic acid and protein sequence analysis: a practical approach (ed. Bishop MJ, Rawlings CJ), pp 173-217. IRL Press, Oxford, United Kingdom.

Staskus KA, Sun R, Miller G, Racz P, Jaslowski A, Metroka C, Brett-Smith H, Haase AT. 1999. Cellular tropism and viral inteleukin-6 expression distinguish human herpesvirus 8 involvement in Kaposi's sarcoma, primary effusion lymphoma, and multicentric Castleman's disease. Journal of Virology 73, 4181-4187.

Staskus KA, Zhong W, Gebhard K, Herndier B, Wang H, Renne R, Beneke J, Pudney J, Anderson DJ, Ganem D, Haase AT. Haase AT. 1997. Kaposi's

sarcoma-associated herpesvirus gene expression in endothelial (spindle) tumor cells. Journal of Virology 71, 715-719.

Steven AC, Spear PG. 1997. Herpesvirus capsid assembly and envelopment. In: Structural Biology of Viruses (ed. Chiu W, Burnett RM, Barcea R), pp. 312-351. Oxford University Press, New York.

Strauchen JA, Hauser D, Burnstein D, Jimenez D, Moore PS, Chang Y. 1996. Body cavity-based malignant lymphoma containing Kaposi's sarcoma-associated herpesvirus in an HIV-negative man with previous Kaposi's sarcoma. Annals of Internal Medicine 125, 822-825.

Stürzl M, Blasig C, Schreier A, Neipel F, Hohenadl C, Cornali E, Ascherl G, Essner S, Brockmeyer NH, Ekman M, Kaaya EE, Tsahachler E, Biberfeld P. 1997a. Expression of HHV-8 latency-associated T0.7 RNA in spindle and endothelia cells of AIDS-associated, classical and African Kaposi's sarcoma (KS). International Journal of Cancer 72, 68-71.

Stürzl M, Schreier A, Blasig C, Zietz Ch, Ensoli B, Browning PJ, Neipel F, Biberfeld P. 1997b. Human herpesvirus 8 (HHV-8) gene expression in Kaposi's sarcoma (KS) primary lesions: an in situ hybridisation study. Journal of Molecular Medicine 75, B185 (abstract 117).

Su I-J, Hsu Y-S, Chang Y-C, Wang I-W. 1995. Herpesvirus-like DNA sequence in Kaposi's sarcoma from AIDS and non-AIDS patients in Taiwan. Lancet 345, 722-723.

Sun R, Lin S-F, Gradoville L, Miller G. 1996. Polyadenylated nuclear RNA encoded by Kaposi's sarcoma-associated herpesvirus. Proceedings of the National Academy of Sciences, USA 93, 11883-11888.

Sun R, Lin SF, Gradoville L, Yuan Y, Zhu F, Miller G. 1998. A viral gene that activates lytic cycle expression of Kaposi's sarcoma-associated herpesvirus. Proceedings of the National Academy of Sciences, USA 95, 10866-10871.

Sun R, Lin SF, Staskus K, Gradoville L, Grogan E, Haase A, Miller G. 1999. Kinetics of Kaposi's sarcoma-associated herpesvirus gene expression. Journal of Virology 73, 2232-2242.

Szilagyi JF, Berriman J. 1994. Herpes simplex virus L particles contain spherical membrane-enclosed inclusion vesicles. Journal of General Virology 75, 1749-1753.

Tappero JW, Conant MA, Wolfe SF, Berger TG. 1993. Kaposi's sarcoma. Epidemiology, pathogenesis, histology, clinical spectrum, staging criteria and therapy. Journal of American Academy of Dermatology 28, 371-395.

Tarn W, Steitz J. 1996. A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro. Cell 84, 801-811.

Taylor JF, Templeton AC, Vogel CL, Ziegler JL, Kyalwazi SK. 1971. Kaposi's sarcoma in Uganda: a clinico-pathological study. International Journal of Cancer 8, 122-135.

Thomas JA, Brookes LA, McGowan I, Weller I, Crawford DH. 1996. HHV8 DNA in normal gastrointestinal mucosa from HIV seropositive people. Lancet 347, 1337-1338.

Thome M, Schneider P, Hofmann K, Fickenscher H, Meini E, Neipel F, Mattmann C, Burns K, Bodmer JL, Schroter M, Scaffidl C, Krammer PH, Peter ME, Tschopp J. 1997. Viral FLICE-inhibitory proteins (FLIPs) prevent apoptosis induced by death receptors. Nature 386, 517-521.

Tomita Y, Naka N, Aozasa K, Cesarman E, Knowles DM. 1996. Absence of Kaposi's sarcoma-associated herpesvirus like DNA sequences (KSHV) in angiosarcomas developing in body-cavity and other sites. International Journal of Cancer 66, 141-142.

Urassa WK, Kaaya EE, Kitinya SN, Lema LN, Amir H, Luande J, Biberfeld G, Mhalu FS, Biberfeld P. 1998. Immunologic profile of endemic and epidemic Kaposi's sarcoma patients in Dar-es-Salaam, Tanzania. International Journal of Molecular Medicine 1, 979-982.

Uthman A, Brna C, Weninger W, Tschachler E. 1996. No HHV8 in non-Kaposi's sarcoma mucocutaneous lesions from immunodeficient HIV-positive patients. Lancet 347, 1700-1701.

van Santen VL. 1993. Characterization of a bovine herpesvirus 4 immediate-early RNA encoding a homolog of the Epstein-Barr virus R transactivator. Journal of Virology 67, 773-784.

Vieira JD, Hunag L, Koelle DM, Corey L. 1997. Transmissible Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) in saliva of men with a history of Kaposi's sarcoma. J. Virol. 71, 7083-7087.

Virgin HT, Latreille P, Wamsley P, Hallsworth K, Weck KE, Dal CA, Speck SH. 1997. Complete sequence and genomic analysis of murine gammaherpesvirus 68. Journal of Virology 71, 5894-5904.

Viviano E, Vitale F, Ajello F, Perna AM, Villafrate MR, Bonura F, Arico M, Mazzola G, Romano N. 1997. Human herpesvirus type 8 DNA sequences in biological samples of HIV-positive and negative individuals in Sicily. AIDS 11, 607-612.

Wabinga HR, Parkin DM, Wabwire-Mangen F, Mugerwa JW. 1993. Cancer in Kampala, Uganda, in 1989-91: changes in incidence in the era of AIDS. International Journal of Cancer 54, 26-36.

Walter PR, Philippe E, Nguemby-Mbina C, Chamlian A. 1984. Kaposi's sarcoma: presence of herpes-type virus particles in a tumor specimen. Human Pathology 15, 1145-1146.

Whitby D, Boshoff C, Luppi M, Torelli G. 1997. Kaposi's sarcoma-associated herpesvirus and multiple myeloma. Science 278, 1971-1972.

Whitby D, Howard MR, Tenant-Flowers M, Brink NS, Copas A, Boshoff C, Hatziouannou T, Suggett FEA, Aldam DM, Denton AS, Miller RF, Weller IVD, Weiss RA, Teddar RS, Schulz TF. 1995. Detection of Kaposi's sarcoma-associated herpesvirus (KSHV) in peripheral blood of HIV-infected individuals predicts progression to Kaposi's sarcoma. Lancet 346, 799-802.

Whitby D, Luppi M, Barozzi P, Boshoff C, Weiss RA, Torelli G. 1998. Human herpesvirus 8 seroprevalence in blood donors and patients with lymphoma from different regions of Italy. Journal of the National Cancer Institute 90, 395-397.

Whitehouse A, Carr IM, Griffiths JC, Meredith DM. 1997. The herpesvirus saimiri ORF50 gene, encoding a transcriptional activator homologous to the Epstein-Barr virus R protein, is transcribed from two distinct promoters of different temporal phases. Journal of Virology 71, 2550-2554.

Whitman JE, Lapps W, Chatlynne L, Masood R, Gill P, Kaplan M, Friedman-Kien A, Said JW, Koeffler P, Ablashi DV. 1997. Human herpesvirus 8: virus production, morphogenesis, infectivity, and seroprevalence. Journal of Human Virology 1, 150.

Wu L, Lo P, Yu X, Stoops JK, Foghani B, Zhou ZH. 2000a. Three-dimensional structure of the human herpesvirus 8 capsid. Journal of Virology 74, 9646-9654.

Wu TT, Usherwood EJ, Stewart JP, Nash AA, Sun R. 2000b. Rta of murine gammaherpesvirus 68 reactivates the complete lytic cycle from latency. Journal of Virology 74, 3659-3667.

Zhong W, Ganem D. 1997. Characterization of ribonucleoprotein complexes containing an abundant polyadenylated nuclear RNA encoded by Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8). Journal of Virology 71, 1207-1212.

Zhong W, Wang H, Herndier B, Ganem D. 1996. Restricted expression of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) genes in Kaposi's sarcoma. Proceedings of the National Academy of Sciences, USA 93, 6641-6646.

Zhu L, Wang R, Sweat A, Goldstein E, Horvat R, Chandran B. 1999. Comparison of human sera reactivities in immunoblots with recombinant human herpesvirus (HHV)-8 proteins associated with the latent (ORF73) and lytic (ORFs 65, K8.1A, and K8.1B) replicative cycles and in immunofluorescence assays with HHV-8-infected BCBL-1 cells. Virology 256, 381-392.

Ziegler JL, Katongole-Mbidde E. 1996. Kaposi's sarcoma in childhood: an analysis of 100 cases from Uganda and relationship to HIV infection. International Journal of Cancer 65, 200-203.

Zimring JC, Goodbourn S, Offermann MK 1998. HHV8 encodes an IRF homologue that represses IRF-1 mediated transcription. Journal of Virology 72, 701-707.

Zong J-C, Ciufo D, Alcendor DJ, Wan X, Nicholas J, Browning PJ, Rady PL, Trying SK, Orenstein JM, Rabkin CS, Su I-J, Powell KF, Croxson M, Foreman KE, Nickoloff BJ, Alkan S, Hayward GS 1999. High-level variability in the ORF-K1 membrane protein gene at the left end of the Kaposi's sarcoma-associated herpesvirus genome defines four major virus subtypes and multiple variants or clades in different human populations. Journal of Virology 73, 4156-4170.

Zong J-C, Metroka C, Reitz MS, Nicholas J, Hayward GS. 1997. Strain variability among Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8): evidence that a large cohort of United States AIDS patients may have been infected by a single common isolate. Journal of Virology 71, 2505-2511.