HUMAN MUTATION 25:396–409 (2005)

RESEARCH ARTICLE

Conservation of the *RB1* Gene in Human and Primates

Theru A. Sivakumaran,[1] Peidong Shen,[2] Dennis P. Wall,[3] Bao H. Do,[2] Kiran Kucheria,[1] and Peter J. Oefner[2]*

[1]*Division of Genetics, Department of Anatomy, All India Institute of Medical Sciences, New Delhi, India;* [2]*Genome Technology Center, Stanford University School of Medicine, Palo Alto, California;* [3]*Department of Systems Biology, Harvard Medical School, Boston, Massachusetts*

Communicated by Georgia Chenevix-Trench

Mutations in the *RB1* gene are associated with retinoblastoma, which has served as an important model for understanding hereditary predisposition to cancer. Despite the great scrutiny that *RB1* has enjoyed as the prototypical tumor suppressor gene, it has never been the object of a comprehensive survey of sequence variation in diverse human populations and primates. Therefore, we analyzed the coding (2,787 bp) and adjacent intronic and untranslated (7,313 bp) sequences of *RB1* in 137 individuals from a wide range of ethnicities, including 19 Asian Indian hereditary retinoblastoma cases, and five primate species. Aside from nine apparently disease-associated mutations, 52 variants were identified. They included six singleton, coding variants that comprised five amino acid replacements and one silent site. Nucleotide diversity of the coding region ($\pi = 0.0763 \pm 1.35 \times 10^{-4}$) was 52 times lower than that of the noncoding regions ($\pi = 3.93 \pm 5.26 \times 10^{-4}$), indicative of significant sequence conservation. The occurrence of purifying selection was corroborated by phylogeny-based maximum likelihood analysis of the *RB1* sequences of human and five primates, which yielded an estimated ratio of replacement to silent substitutions (ω) of 0.095 across all lineages. *RB1* displayed extensive linkage disequilibrium over 174 kb, and only four unique recombination events, two in Africa and one each in Europe and Southwest Asia, were observed. Using a parsimony approach, 15 haplotypes could be inferred. Ten were found in Africa, though only 12.4% of the 274 chromosomes screened were of African origin. In non-Africans, a single haplotype accounted for from 63 to 84% of all chromosomes, most likely the consequence of natural selection and a significant bottleneck in effective population size during the colonization of the non-African continents. Hum Mutat 25:396–409, 2005. © **2005 Wiley-Liss, Inc.**

KEY WORDS: retinoblastoma; RB1; SNP; haplotype; nucleotide diversity; HKA test; maximum likelihood ratio

DATABASES:
RB1 – OMIM: 180200; GenBank: M15400.1 (mRNA), L11910.1 (genomic), NP_000312 (protein), AF336015 - AF336040 (mRNA, chimpanzee); HGMD: RB1
http://insertion.stanford.edu/melt.html (DHPLC Melt program)
www.ncbi.nlm.nih.gov/SNP/ (dbSNP)

INTRODUCTION

The knowledge of global sequence diversity is of importance in understanding the history and organization of the complex links between genotypic and phenotypic variation in traits that are determinants of human disease susceptibility and resistance [Lander, 1996]. The common types of sequence variation in human include SNPs, small insertions and deletions, and variations in repeat motifs. SNPs occur on average once every 1,000–2,000 bases [Altshuler et al., 2000; Shen et al., 2002; Wang et al., 1998]. SNPs can serve as genetic markers in identifying disease genes in families by linkage studies, linkage disequilibrium in isolated populations, and association analysis of patients and controls [Collins et al., 1997; Kruglyak, 1997]. Although large-scale population studies of DNA sequence variation have become available for an increasing number of genes [Cambien et al., 1999; Cargill et al., 1999; Glatt et al., 2001; Halushka et al., 1999; Harris and Hey, 1999; Nickerson et al., 1998; Passarino et al., 2001; Rieder et al., 1999; Sale et al., 2002; Shen et al., 2000; Tiret et al., 2002; Twells et al., 2003], only a few such studies have been published on genes predisposing to cancer [Thorstenson et al., 2001; Wagner et al., 1999].

Received 25 May 2004; accepted revised manuscript 13 December 2004.

*Correspondence to: Peter J. Oefner, Stanford Genome Technology Center, 855 California Avenue, Palo Alto, CA 94304.
E-mail: oefner@genome.stanford.edu

Grant sponsors: International Union Against Cancer (UICC); International Cancer Technology Transfer Fellowship; Grant sponsor: National Institutes of Health; Grant number: GM28428.

Current address of T.A. Sivakumaran: Fellow, Genetics Training Program, Harvard Medical School, HMS-New Research Building 160, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115.

The Supplementary Material referred to in this article can be accessed at http://www.interscience.wiley.com/jpages/1059-7794/suppmat.

DOI 10.1002/humu.20154
Published online in Wiley InterScience (www.interscience.wiley.com).

The human *RB1* (MIM# 180200; GenBank M15400.1) gene is an important case study for mutational analysis. *RB1* was the first tumor suppressor gene isolated and is commonly associated with an embryonic neoplasm of retinal origin called retinoblastoma [Fung et al., 1987; Goddard et al., 1988; Lee et al., 1987a]. The gene is located at 13q14, spans about 180 kb, and comprises 27 exons. The *RB1* gene transcribes 4.7 kb of mRNA (GenBank M15400.1), which is expressed in a wide variety of tissues. The open reading frame of the *RB1* gene encodes a 928–amino acid nucleophosphoprotein of 110 KD (GenBank NP_000312) [Lee et al., 1987b]. The RB protein contains several functional domains [Harbour, 2001]. The largest domain is the pocket domain (amino acids 379–792). It contains two highly conserved regions called the A and B boxes, which have to interact first to attain an active conformation for binding to several viral oncoproteins and cellular regulatory proteins that contain the amino acid motif LXCXE and play a critical role in controlling cell growth. In addition, the pocket domain of the RB protein interacts with E2F transcription factors (e.g., MIM# 189971) resulting in repression of the transcription of genes required for the G1 to S phase transition. The carboxy-terminal region (amino acids 793–928) contains a nuclear localizing signal, a second binding site for E2F, binding sites for MDM2 (MIM# 164785) and c-Abl (MIM# 189980), seven consensus cyclin-dependent kinase (CDK) phosphorylation sites, and docking sites for CDK4 (MIM# 123829)/CDK6 (MIM# 603368). Six additional consensus CDK phosphorylation sites are located in the least understood amino-terminal region (amino acids 1–378), which may play a role in regulating RB protein in the cell cycle. Inactivation of both copies of the *RB1* gene results in retinoblastoma [Knudson, 1971]. The spectrum of causative mutations ranges from large deletions to single base substitutions, with over 400 mutations in the *RB1* gene having been reported to date [Alonso et al., 2001; Ata-ur-Rasheed et al., 2002; Blanquet et al., 1993, 1994, 1995; Cowell and Cragg, 1996; Cowell et al., 1994; Dalamon et al., 2001; Houdayer et al., 2004; Jakubowska et al., 2001; Kiran et al., 2003; Kloss et al., 1991; Kumaramanickavel et al., 2003; Liu et al., 1995; Lohmann et al., 1992, 1994a, 1994b, 1996, 1997; Richter et al., 2003; Yilmaz et al., 1998] (www.d-lohmann.de/Rb/mutations.html). The majority of these mutations disrupt protein function, i.e., nonsense and frameshift mutations, which account for about 76% of the mutations, followed by splice site mutations (15%), missense mutations (6%), in-frame deletions (2%), and mutations in the promoter region (1%). Apart from these disease-causing mutations, about 28 DNA sequence variations/polymorphisms have been reported in a limited number of populations [Lohmann, 1999].

This study reports not only apparently disease-associated mutations in *RB1* in 19 hereditary retinoblastoma cases from India, but also general sequence variation in the coding, flanking noncoding and promoter regions of *RB1* in 94 representatives from the five continents and 24 additional controls from the Indian subcontinent. Analysis of sequence variation in *RB1* was extended to its orthologs in chimpanzee, gorilla, orangutan, rhesus monkey, and black-handed spider monkey. Sequence variants were used to infer haplotypes and to determine their frequency and geographic distribution.

MATERIALS AND METHODS

Retinoblastoma Patients

A total of 19 patients with confirmed diagnosis of retinoblastoma at the Dr. Rajendra Prasad Center for Ophthalmic Science, All India Institute of Medical Sciences, New Delhi, India, were included in this study after approval by the institutional review board.

Human Diversity and Primate DNA Samples

A total of 118 control individuals from worldwide populations were screened for sequence variants in the *RB1* gene. All samples were collected according to approved human subject protocols. The populations surveyed were as follows: 17 Africans (three Zaire Pygmies, two Car Pygmies, four Lissongos, two Ethiopian Jews, one Zulu, one Sudanese, one Ghanian, one Bushman, one Namibian San, and one Tsumkwe San), 10 Native American Indians (one Karitiana, one Surui, two Mayans, two Columbia Indians, one Quechua, one Waorani, one Pima, and one Navaho), 64 Asians (two Bedouine, one Hmong, one Tamil, two Druze, two Palestinians, two Sepharadim, one Yakut, one Iranian, one Pakistani, one Pathan, two Sindhi, two Brushaski, one Baloochi, one Brahui, one Makran, one Hazara, one Kalahkafir, 24 North Indians from the states of Punjab, Haryana, Utter Pradesh, and New Delhi, three Cambodians, one Taiwanese, one Atayal, one Ami, six Han Chinese, four Japanese, and one Korean), 19 Europeans (two Finns, one Adygei, one Russian, one Caucasus, one Georgian, one Greek, two Ashkenazi, two French, and one Utah from the Centre d'Etude des Polymorphism Human (CEPH) collection, one Amish, two English, two Italians, and one Dane) and eight Oceanians (one Micronesian, two Australian Aborigines, two Melanesians, two New Guineans, and one Samoan).

PCR primers designed from the human *RB1* sequence (Supplementary Table S1, available online at http://www.interscience.wiley.com/jpages/1059-7794/suppmat) were used to amplify the homologous regions of *RB1* for sequencing in chimpanzee (*Pan troglodytes*), lowland gorilla (*Gorilla gorilla*), orangutan (*Pongo pygmaeus*), old world rhesus monkey (*Macaca mulatta*), and new world black-handed spider monkey (*Ateles geoffroyi*). The samples were obtained from Coriell Cell Repositories (Camden, NJ). The five chimpanzee samples were also analyzed by DHPLC for all *RB1* coding regions except exon 1 and the variants detected were submitted to GenBank (AF336015–AF336040).

PCR Conditions

Genomic DNA was prepared through use of a QIAamp DNA Mini Kit (Qiagen, Valencia, CA). Primers were designed using the complete *RB1* genomic sequence (GenBank L11910.1) reported by Toguchida et al. [1993]. PCR was performed in a 50-μl volume containing 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 50 μM dNTPs, 0.25 μM of each primer, 50 ng of genomic DNA, and 1 unit of Ampli*Taq* Gold (Applied Biosystems, Foster City, CA). The $MgCl_2$ concentration was typically 2.5 mM, except for exon 1 (1.5 mM) and exon 25 (3.5 mM). The PCR cycling regime comprised an initial denaturation step at 95°C for 10 min. This was typically followed by 14 "touchdown PCR" cycles of 94°C for 20 s, 63°C for 1 min (decreasing by 0.5°C/cycle), and 72°C for 1 min; 20 cycles of 94°C for 20 s, 56°C for 45 s, and 72°C for 45 s; and 1 cycle of 72°C for 7 min. (The only exception was exon 1, for which no touchdown, but a constant annealing temperature of 64°C was used for the first 14 cycles).

DHPLC

DHPLC was carried out on automated HPLC instrumentation equipped with a DNA Sep column (Transgenomic, Omaha, NE). Crude PCR products were subjected to a 3-min denaturing step at 95°C, followed by gradual reannealing over 30 min by lowering the temperature from 95–65°C. Amplicons were eluted with a linear acetonitrile (J.T. Baker, Phillpsburg, NJ) gradient at a flow rate of 0.9 ml/min. The start- and end-points of the gradient were adjusted according to the size of the PCR products using an algorithm provided by the WAVE maker system control software (Transgenomic, San Jose, CA). Generally, analysis time was 8 min, including column regeneration and reequilibration to the starting

conditions. The temperature required for successful resolution of heteroduplex molecules was determined by use of the DHPLC melting algorithm [Xiao and Oefner, 2001]. Details about the temperature conditions and algorithm can be obtained at http://insertion.stanford.edu/melt.html. The temperature and gradient conditions at which PCR products were analyzed are given in Supplementary Table S1. The 5′ untranslated region (5′UTR), with the exception of nt 1065–1143 (GenBank L11910.1) that contain a run of 23 As not amenable to sequence analysis, flanking intronic sequences and all exons except exon 8, which is not translated, were subjected to DHPLC analysis.

Direct Sequencing

PCR products were purified by solid-phase extraction and were bidirectionally sequenced with the Amersham (Piscataway, NJ) Thermosequenase II Dye Deoxy Terminator Cycle Sequencing Kit according to the manufacturer's instructions. Samples were analyzed with an Applied Biosystems 377 sequencer.

Data Analysis

Nucleotide diversity (π) was calculated from the equation

$$\hat{\pi} = \sum_{i<j} \pi_{ij} \neq n_c$$

where π_{ij} is the number of nucleotide differences between the ith and jth DNA sequences and $n_c = n(n-1)/2$ [Nei, 1987].

We tested for the historic occurrence of positive, directional selection using methods described in Yang [1998] and Nielsen and Yang [1998]. In these tests, maximum likelihood ratios of nonsynonymous to synonymous mutations (ω, omega) exceeding 1 are considered evidence for the existence of positive selection. We conducted both lineage- and site-specific tests. For the lineage-specific tests, we used a model in which all lineages have the same ω (hereafter referred to as M0) and compared that with a model in which ω is estimated for each lineage (hereafter referred to as M1). For the site-specific tests, we compared M0 with a discrete model that allows for site-to-site heterogeneity in positive selection (hereafter referred to as M3). To better understand the intensity of purifying selection, we tested models that fix omega at values between 0 and 1 incrementally against M0. This procedure allowed us to determine the range of omega values that are consistent with the null model, M0 (that is where twice the difference in log likelihoods ($2\Delta\ln L$) of the two models does not differ by more than the χ^2 critical value of 3.84 for a one-tailed test with one degree of freedom). Finally, to test for the action of selection among amino acid sites within a specific lineage, we compared a model that allows for heterogeneity in ω among sites but not among lineages, with a model that allows for variation in ω along a predefined lineage (as in Yang and Nielsen [2002]). We assumed the following phylogeny ((((mus,orangutan),gorilla), bonobo),chimp,human) for our *RB1* tests. However, results of our analyses were robust to minor fluctuations in the tree.

Tests of the occurrence of directional selection acting on the *RB1* gene also included the Hudson-Kreitman-Aguadé (HKA) test [Hudson et al., 1987]. The HKA test compares levels of intraspecies polymorphism at different loci and interspecies sequence differences at these loci, with expectations (under a neutral evolution model) of an infinite number of sites. A neutral model predicts that the ratio between levels of intraspecies polymorphism and interspecies divergence should be the same for different genes, because both are a function of the neutral mutation rate. Selection, in contrast, leads to an uncoupling of levels of polymorphism and divergence. We also asked whether nonsynonymous and synonymous mutation rates differed within and among species lineages. As in Hasegawa et al. [1998], we compared a model in which intraspecific and interspecific lineages have the same ω with a model that assumes a different ω for intraspecific lineages. We applied this test to the coding regions of both *RB1* and *ATM* using the unrooted trees (human2, ((orang2,orang1),(chimp1,chimp2)),human1) and (chimp2,(((macaca,orangutan),gorilla)(human1,human2))chimp1), respectively.

Log-linear modeling [Huttley and Wilson, 2000] was applied to test the level of pairwise linkage disequilibrium in the *RB1* gene region between African and non-African populations for the polymorphic sites with a minor allele frequency $\geq 5\%$. In the analysis, a series of multiplicative models was constructed, each with a different disequilibrium term. The fit of a model was measured as the likelihood-ratio test statistic or deviance from the full model. The difference in deviances (ΔD) for two models is distributed approximately as χ^2 and can be used to test the significance with and without the particular disequilibrium term. The analysis was implemented in the statistical programming language R (version 1.9.1) through use of the package hwde (contributed by J. Maindonald; available from the R Project [www.r-project.org] for Statistical Computing Web site). Conventional measures of linkage disequilibrium, namely $|D'|$ and r^2, were also calculated for each pair of SNPs in *RB1* that had a minor allele frequency $\geq 5\%$ [Hartl and Clark, 1997]. Significance levels were determined by the χ^2 statistic for the corresponding 2×2 table (one degree of freedom).

Inference of Haplotypes from SNPs

Haplotypes were inferred by a maximum-parsimony approach as described previously [Jin et al., 1999] for each individual for which complete genotyping data had been obtained at the 13 polymorphic sites with a minor allele frequency $>2\%$. Polymorphic sites from a group of haplotypes or sequences were considered congruent if they could be accommodated by the same phylogenetic topology. Polymorphic sites were considered recurrent if linkage to flanking markers could not be explained by a single recombination event. In the construction of the tree by maximum parsimony, haplotypes of each individual with multiple polymorphisms were inferred, assuming the least number of mutational steps and any incongruence between loci indicating recombination or recurrent mutation.

RESULTS

Apparently Disease-Associated *RB1* Gene Mutations

The peripheral blood leukocyte DNAs from three familial bilateral cases, 15 sporadic bilateral cases, and a sporadic unilateral case with pineoblastoma were analyzed to identify the causative *RB1* germline mutations. Of the 19 patients studied, one familial bilateral case (151/96) and eight sporadic bilateral cases were found to harbor apparently pathogenic mutations, including two nonsense mutations in exons 14 and 23, four splice site mutations in introns 1, 3, 12, and 17, one frameshift mutation in exon 4, one in-frame deletion in exon 20, and a missense mutation in exon 17 (Table 1). The DNA samples from the parents of these patients were not available for genotyping. Only three of the mutations had been reported previously. They included the splice site mutation *c.1215+1G>A* (GenBank M15400.1) [Dunn et al., 1989; Houdayer et al., 2004; Kiran et al., 2003; Klutz et al., 1999; Lohmann et al., 1996, 1997; Richter et al., 2003; Yilmaz et al., 1998] and the two nonsense mutations *c.1333C>T* [Choy et al., 2002; Houdayer et al., 2004; Kiran et al., 2003; Klutz et al., 1999; Lohmann et al., 1996, 1997; Richter et al., 2003; Yandell et al., 1989] and *c.2359C>T* [Alonso et al., 2001; Cowell et al., 1994; Houdayer et al., 2004; Lohmann et al., 1996, 1997; Richter et al., 2003; Yandell et al., 1989; Yilmaz et al., 1998; Zhang et al., 1997]. A homozygous 2-bp deletion (*c.380+17_380+18delAA*) in intron 3 and a heterozygous transversion (*c.607+200T>A*) in intron 6, respectively, were detected only once each in the 19 cases of bilateral retinoblastoma (Table 2). Their functional significance, if any, is unknown.

TABLE 1. Apparently Disease-Causing Mutations Found in Nine Indian Bilateral Retinoblastoma Patients

Patients	Genomic position[a] (location within gene)	Nucleotide change[b]	Predicted protein or splicing alteration	DHPLC temperature (°C)	Reference
108/98	2197 (intron 1)	c.137+1G>T	Donor splice site	67	
234/97	41925 (intron 3)	c.381–1G>C	Acceptor splice site	53	
157/97	41989 (exon 4)	c.444–445insAT	p.M149fsX153[b]	53	e
39/98	70330 (intron 12)	c.1215+1G>A	Donor splice site	53	f
86/97	76430 (exon 14)	c.1333C>T	p.R445X[c]	61	
203/98	78273 (exon 17)	c.1689G>T	p.W563[c,d]	51	
256/96	78280 (intron 17)	c.1695+1G>T	Donor splice site	55	
151/96	156695 (exon 20)	c.1963_1965delTAT	p.Tyr655del[c]	58	
197/97	162237 (exon 23)	c.2359C>T	p.R787X[c]	56	g

[a]Genomic position in GenBank L11910.1.
[b]According to GenBank MI5400.1.
[c]According to GenBank NP_000312.1.
[d]Nonconservative amino acid change.
[e]Dunn et al. [1989]; Houdayer et al. [2004]; Kiran et al. [2003]; Klutz et al. [1999]; Lohmann et al. [1996, 1997]; Richter et al. [2003]; Yilmaz et al. [1998].
[f]Choy et al. [2002]; Houdayer et al. [2004]; Kiran et al. [2003]; Klutz et al. [1999]; Lohmann et al. [1996, 1997]; Richter et al. [2003]; Yandell and Dryja [1989].
[g]Alonso et al. [2001]; Cowell et al. [1994]; Houdayer et al. [2004]; Lohmann et al. [1996, 1997]; Richter et al. [2003]; Yandell and Dryja [1989]; Yilmaz et al. [1998]; Zhang et al. [1997].

Sequence Variants Detected in Worldwide Populations

Aside from the two aforementioned variants of unknown functional significance that had been identified in 2 of 19 retinoblastoma patients, a total of 50 sequence variants were detected in 10,100 bp of coding, 5′ untranslated, and flanking noncoding regions of the *RB1* gene in 118 control individuals screened by DHPLC and direct sequencing (Table 2). Only seven had been reported previously [Blanquet et al., 1995; Lohmann et al., 1997; Mateu et al., 1997; Schubert and Hansen, 1996; Yandell and Dryja, 1989]. The most frequent form of DNA variation was single-base substitutions (90%), while small deletions and insertions accounted for 6 and 4%, respectively. Among the single-base substitutions, transitions were more prevalent (71%) than transversions (29%). Most of the screened sequence was noncoding (7,313 bp) and as expected the majority of the variable sites (N=44) were found in the noncoding region (1/166). In the coding region, only 6 out of 2,787 nucleotides were variable (1/465). Five of the coding variants were nonsynonymous (83%), only one of which resulted in a nonconservative amino acid change (*p.E746G*; GenBank NP_000312.1), and one was silent (17%). All coding variants were observed once in this study. Only *p.E137D* had been reported previously in a patient with isolated unilateral retinoblastoma [Lohmann, et al., 1997]. Although it was the only potentially pathogenic mutation in both the tumor and constitutional cells of the patient, it remains unclear whether this alteration is a predisposing mutation or a neutral variant. Three each of the coding variants were found in 1,788 nondegenerate and 643 two-fold degenerate nucleotides, respectively, while no variant was detected in 356 four-fold degenerate bases. Four of those variants, including the single silent one, were observed in the 24 North Indian controls. The remaining two had been detected in an Ethiopian Jew and an Iranian individual, respectively.

Frequency and Global Distribution of Sequence Variants

A total of 34 sequence variants (68%) were detected solely on one continent, of which 23 (46%) were observed once. Of the 11 continent-specific polymorphisms observed more than once, only four were specific for an ethnic group, namely Pygmy. Interestingly, those four variants were observed in Pygmies from both Zaire and the Central African Republic and, therefore, must predate their split into two isolated societies. Of the 23 variants observed only once, 11, 9, 2, and 1 had been detected in Africa, Asia, Oceania, and America, respectively. Of the remaining 16 polymorphisms, three were observed on two continents: *c.1–732A>C* [GenBank M15400.1] in Southwest Asia and Southeast Europe (Georgian Republic); *c.540–23_540-22insT* on the Indian subcontinent and in Oceania; and *c.718+138G>A* in Africa and Asia. Four polymorphisms (*c.264+75T>C*, *c.381–152A>G*, *c.1127+74C>T*, and *c.1815–104A>G*) were observed in Africa, along the southern coast of Asia, and in Oceania, representing most likely the remnants of the first dispersal of anatomically modern humans out of Africa [Jin et al., 1999]. A fifth tricontinental polymorphism had been detected in all parts of Asia, Europe, and America. Eight polymorphisms were observed on four continents: six on every continent with the exception of Europe, while two were found everywhere other than in Oceania.

A total of 68% (34/50) of the polymorphisms discovered in the 118 control individuals were found in Africa. Thirteen of these polymorphisms were also detected in other continents. Among those was *c.718+138G>A* in intron 7; it had been found in three Lissongos as well as in an individual from the Makran coast of Pakistan. The same individual had been already observed to be the only non-African to share *p.D126E* of the *ATM* (MIM# 607585; GenBank: HSU82828) gene with Africans [Thorstenson et al., 2001]. This had been considered as potential evidence for a relatively recent introduction of alleles from Africa through trade relationships.

NUCLEOTIDE DIVERSITY

Nucleotide diversity of the coding region of *RB1* was very low ($\pi=0.0763\pm1.35\times10^{-4}$) in the 94 representatives of worldwide populations routinely studied in our laboratory. This estimate is about 10-fold lower than those obtained for *ATM* ($\pi=0.71\times10^{-4}$) [Thorstenson et al., 2001] and *CSTB* (MIM# 601145; GenBank: J00314) ($\pi=0.74\times10^{-4}$) [Sale et al., 2002], which had yielded in our hands the lowest nucleotide diversity

TABLE 2. Sequence Variants Found in *RB1*, DHPLC Column Temperature at Which Polymorphism Was Observed, and Number of Heterozygotes Observed in Different Geographic Locations

Genomic position[a] (location in gene)	Polymorphism[b]	Amino acid change[c]	DHPLC Temperature (°C)	No. of heterozygous individuals in Africa (n=17)	West Asia (n=9)	Indian Subcontinent (n=55)	East Asia (n=19)	Oceania (n=8)	Americas (n=10)	Europe (n=19)	Total (n=137)	References
195 (5′ UTR)	c.1-1865_1-1864insCTGATA		58	4							4	
198 (5′ UTR)	c.1-1862delC		58	4							4	
371 (5′ UTR)	c.1-1689C>A		58	1	3	19	6		3	5	37	
605 (5′ UTR)	c.1-1455T>C		58	1							1	
914 (5′ UTR)	c.1-1146C>T[d]		59	2							2	
1023 (5′ UTR)	c.1-11037G>A		59		1						1	
1219 (5′ UTR)	c.1-841C>T[d]		61			1					1	
1282 (5′ UTR)	c.1-778C>T[d]		61	1							1	
1328 (5′ UTR)	c.1-732A>C		61		1	5				1	7	
1409 (5′ UTR)	c.1-641A>G		61	1							1	
1601 (5′ UTR)	c.1-459C>T[a]		62			1					1	
5388 (intron 1)	c.1-38-36_138-31delTATAAG		53					1			1	
5625 (intron 2)	c.264+75T>C[d]		53	2		2		2			6	Blanquet et al. [1995]
39573 (intron 3)	c.380+12T>C[d]		55	2							2	Blanquet et al. [1995]
39578 (intron 3)	c.380+17_380+18delAA		–[f]			1					1	
39606 (intron 3)	c.380+45C>T[d]		55	1	3	19	6		3	5	37	Blanquet et al. [1995]
41774 (intron 3)	c.381-152G>A		53	10	1	4	3	4			22	
41942 (exon 4)	c.397A>C	p.N133H	53	1							1	
41956 (exon 4)	c.411A>T	p.E137D	53			1					1	Lohmann et al. [1997]
42068 (intron 4)	c.500+23G>T		53	2	1	5		1	1		10	
42091 (intron 4)	c.500+46A>G		53	1							1	
44591 (intron 4)	c.501-77G>A		50	7	1	4		1	1		14	
44630 (intron 4)	c.501-38G>A		50	4							4	
45776 (intron 5)	c.540-23_c.540-22insT		53			2		2			4	
46065 (intron 6)	c.607+200T>A		53			1					1	
56882 (intron 6)	c.608-31C>T		54						1		1	
56989 (intron 7)	c.718+26T>C		54				1				1	
56998 (intron 7)	c.718+35A>T		54	1							1	
57101 (intron 7)	c.718+138G>A		54	3		1					4	
59473 (intron 7)	c.719-178T>C		51	3							3	
61670 (intron 8)	c.862-60T>G		54					1			1	
61894 (intron 9)	c.939+87T>C[d]		54	1							1	
64497 (intron 10)	c.1049+58G>A		53	1							1	
65515 (intron 11)	c.1127+74C>T[d]		55	7		4	3	4			18	
65541 (intron 11)	c.1127+100T>C[d]		55	1							1	
65614 (intron 11)	c.1127+173G>A		–[f]	1							1	
70170 (intron 11)	c.1128-72T>G		53	2	1	3		1	1		8	
73724 (intron 12)	c.1216-29A>G		55	2							2	
73937 (intron 13)	c.1332+68T>A		55	3							3	
76678 (intron 14)	c.1389+192delC		56	2							2	
76717 (intron 14)	c.1389+231A>T		56		1	3					4	
76721 (intron 14)	c.1389+235G>C		56	6	2	8	3	5	1		25	
76772 (intron 14)	c.1389+286G>A		56	1							1	
150009 (exon 18)	c.1707A>G	p.L569I	55			1					1	
150098 (exon 18)	c.1796A>G	p.N599S	55		1						1	
153095 (intron 18)	c.1815-113A>G		56	2							2	
153104 (intron 18)	c.1815-104A>G		56	5		4	3	4			16	Schubert and Hansen [1996]

TABLE 2. (continued)

Genomic position[a] (location in gene)	Polymorphism[b]	Amino acid change[c]	DHPLC Temperature (°C)	No. of heterozygous individuals in							References
				Africa (n=17)	West Asia (n=9)	Indian Subcontinent (n=55)	East Asia (n=19)	Oceania (n=8)	Americas (n=10)	Europe (n=19)	Total (n=137)	
156616 (intron 19)	c.1961-77A>G		54	7	1	4		1	1		14	Mateu et al. [1997]
156823 (exon 20)	c.2091C>G	p.D697E	54			1					1	
162022 (exon 22)	c.2237A>G	p.E746G[e]	56			1					1	
173882 (intron 25)	c.2663+33T>C[d]		58	6	1	4		1	1		13	
174351 (intron 25)	c.2664-10A>T		55		3	18	6		3	4	34	Yandell and Dryja [1989]

[a]Genomic position in GenBank L 11910.1.
[b]According to GenBank M15400.1.
[c]According to GenBank NP_000312.1.
[d]DNA variant at *CpG* or *CpNG* site.
[e]Nonconservative amino acid change.
[f]Homozygous base changes detected during genotyping of 39606 C→T and 65515 C>T, respectively, by dye-terminator sequencing.

values thus far in the same set of individuals. Even in comparison to *SLC6A4* (MIM# 182138), the gene with the lowest value of nucleotide diversity reported to date in the literature [Glatt et al., 2001], *RB1* yields a five-fold lower estimate of nucleotide diversity. Overall, nucleotide diversity appears to be 45-fold less compared to the arithmetic mean (3.47×10^{-4}) of 12 other autosomal genes analyzed by DHPLC [Sale et al., 2002], and 65-fold less than the nucleotide diversity value reported in 135,823 coding base pairs in 106 human genes [Cargill et al., 1999].

There was a large difference in *RB1* nucleotide diversity between coding and noncoding regions. The nucleotide diversity of the latter was $3.93 \pm 5.26 \times 10^{-4}$. Whereas most autosomal and Y-chromosomal genes in our hands [Passarino, et al. 2001; Sale et al., 2002; Shen et al., 2000; Thorstenson et al., 2001; Wagner et al., 1999] have similar nucleotide diversity estimates in the protein coding and noncoding regions, with an average ratio of 1:2.3, in *RB1* the ratio is 1:51.5. In comparison, the respective ratios for the only two other autosomal genes with ratios significantly greater than 1:2.5, namely *ATM* and *TUBB* [MIM# 191130], were 1:7.5 [Thorstenson et al., 2001] and 1:6.3 [Sale et al., 2002]. Nucleotide diversity in the noncoding region of *RB1* is only slightly lower than that of other autosomal genes that were resequenced in the same set of individuals (average 8.07×10^{-4}, with a range from 4.00×10^{-4} to 14.56×10^{-4}). Hence, mutation rate in the chromosomal region containing *RB1* is not generally lower. More likely, the low sequence diversity in the coding region is due to selective pressure for maintaining the protein sequence and preferred codons.

When total nucleotide diversity, including both coding and noncoding sequences, was calculated for African and non-African populations separately, the latter had 3.1 times less sequence diversity ($1.85 \pm 2.50 \times 10^{-4}$) than did the African population ($5.73 \pm 4.35 \times 10^{-4}$). The values of nucleotide diversity for non-African and African populations, respectively (with ratios given in parentheses following the diversity values), in *ATM* [Thorstenson et al., 2001], *BRCA2* (MIM# 600185; GenBank: U43746) [Wagner et al., 1999], *WRN* (MIM# 277700; GenBank: AF091214) [Passarino et al., 2001], and *TUBB* [Sale et al., 2002] were as follows: in *ATM*, $2.55 \pm 2.32 \times 10^{-4}$ and $3.70 \pm 2.71 \times 10^{-4}$ (1:1.45); in *BRCA2*, $2.84 \pm 2.73 \times 10^{-4}$ and $3.54 \pm 4.22 \times 10^{-4}$ (1:1.25); in *WRN*, $4.82 \pm 3.11 \times 10^{-4}$ and $7.13 \pm 4.09 \times 10^{-4}$ (1:1.48); and in *TUBB*, $3.01 \pm 3.66 \times 10^{-4}$ and $6.99 \pm 4.28 \times 10^{-4}$ (1:2.32).

Comparison of *RB1* Orthologs

Table 3 shows a comparison of the fixed and polymorphic *RB1* coding and noncoding sequence differences observed between human and primates. The single fixed amino acid change detected between human and chimpanzee in the exons successfully sequenced (exon 1 failed) was *p.M233V* (GenBank NP_000312.1) (*c.697A>G*; GenBank M15400.1). This conservative amino acid replacement change was also observed in orangutan. The five additional fixed amino acid changes observed between human and orangutan (sequencing failed for exons 1) were *p.L110Q* (*c.329T>A*), *p.S114P* (*c.340T>C*), *p.A188S* (*c.562G>T*), *p.N316S* (*c.947A>G*), and *p.V368I* (*c.1102G>A*). With the exception of the two changes in exon 3 (*p.L110Q* and *p.S114P*), these were all conservative amino acid replacements. One of the nonconservative changes, *p.S114P,* was also detected in gorilla. The second fixed difference between human and gorilla (exons 1 failed) was also nonconservative in nature and located in exon 5 (*p.L171F, c.511C>T*). In

TABLE 3. Numbers of Base Pairs Screened, Numbers of Fixed and Polymorphic DNA Sequence Variants Observed in Primates, and Results of the Hudson-Kreitman-Aguadé Test

Region	Coding					Noncoding			Selection tests
		Fixed differences		Polymorphic differences					
Species (no. of chromosomes)	Base pairs analyzed	Replacement	Silent	Replacement	Silent	Base pairs analyzed	Fixed differences	Polymorphic differences	HKA[a] P[c]
Chimpanzee (10)	2646	1	1	2	0	7184	58	8	1.00
Gorilla (4)	2646	2	5	0	0	4681	44	4	0.79
Orangutan (4)	2646	6	21	0	2	4441	94	9	0.114
Rhesus (4)	2527	6	37	0	0	3452	149	8	0.282
Spider monkey (2)	1678	15	30	1	1	2271	174	1	n.a.[d]
Mouse[b] (2)	2766	88	196	n.a.	n.a.	n.a.	n.a.	n.a.	0.550

[a]Hudson-Kreitman-Aguadé test [Hudson et al., 1987]; the entire coding region of *RBI* was compared against the first 71% of the coding region of *ATM* [Thorstenson et al., 2001].
[b]GenBank Accession number: M26391.1.
[c]After Bonferroni correction.
[d]Data for ATM is not available.

conclusion, all fixed amino acid replacement differences between human and great apes were located in the N-terminal region that appears to be important but not completely essential for pRB function and tumor suppression [Harbour, 2001].

Statistical Analysis for the Hypothesis of Selection at *RB1*

The HKA test [Hudson et al., 1987] compares the number of intraspecies (within-human) polymorphisms and interspecies differences at different loci, with the expectation of a model of neutral evolution. Four cross-species comparisons were performed between the *RB1* locus and the first 71% of coding region of *ATM*, for which recent analysis [Thorstenson et al., 2001] had shown absence of significant selective pressure. None of the comparisons were statistically significant ($P<0.05$) after Bonferroni correction. To complement our HKA test, we asked whether nonsynonymous and synonymous mutation rates differed within and among species lineages in *RB1* and *ATM*. As in Hasegawa et al. [1998], we compared a model in which intraspecific and interspecific lineages have the same ratio of nonsynonymous to synonymous mutations (ω) with a model that assumes a different ω for intraspecific lineages. Differences between interspecies and intraspecies mutation rates were not significant.

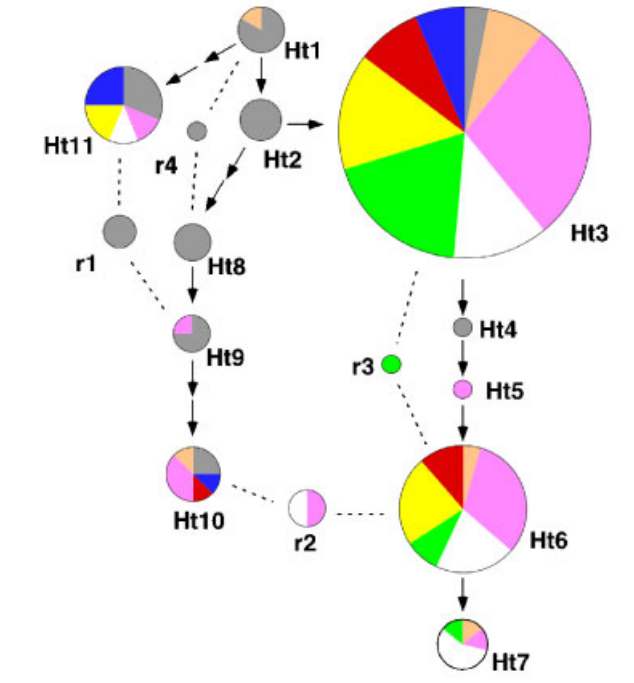
Estimates of positive selection among sites within the coding region of *RB1* (minus exon 1) revealed a small percentage of sites ($\sim$1.2%) with ratios of numbers of nonsynonymous (d_N) over synonymous substitutions (d_S) per site above 3.3 and as high as 4.7. Two sites (codons 188 and 233) were found to be under significant positive selection ($P<0.01$). However, the gene appears to have evolved largely via purifying selection throughout the history of primates. Given that our lineage specific likelihood tests could not reject a null hypothesis of one ω for all lineages, we were able to compare estimates of ω in *RB1* and *ATM* for all lineages. Compared with *ATM*, which has an estimated omega of 0.18 across all lineages, the omega for *RB1* was found to be approximately half as large, at 0.095, suggesting that stronger negative selection has operated on *RB1*. In an attempt to quantify the extent of purifying selection that has operated on *RB1,* we ran a model that assumes selective neutrality in the *RB1* sequences (omega=1). In a likelihood ratio test [Yang, 1998], we found this model to be significantly worse than M0 ($2\Delta\ln L$-340.92; $P\approx 0$), a fact consistent with deviation from selective neutrality but insufficient to quantify the extent of purifying selection. Additional tests using models in which omega is fixed at values incrementally greater than 0 showed that the ratio of nonsynonymous to synonymous mutations in *RB1* is between 0.08 and 0.1, corroborating that this gene is under significant evolutionary constraint. As a final qualitative look at the degree of negative selection operating on *RB1*, i.e., the extent to which the rates of synonymous divergence are less than the neutral mutation rate, we compared the branch lengths estimated from the intron sequences of *RB1* with those estimated from the silent substitution rate of adjacent exons. We found that all of the branch lengths calculated from the silent substitution rate are shorter than the branch lengths calculated using the intron-based and, thus, selectively neutral tree, providing additional support for our hypothesis of purifying selection in *RB1*. The cause of this marked negative selection could be selection for codon usage bias to enhance translational accuracy or efficiency [Bulmer, 1991; Plotkin et al., 2004; Urrutia and Hurst, 2003], but more data are required to adequately test this hypothesis.

Haplotype Analysis and Linkage Disequilibrium

Haplotypes for each individual were inferred by a maximum-parsimony approach, using the genotype data of 13 polymorphic sites with a minor allele frequency $>2\%$. All loci had been detected in noncoding regions of *RB1*. The genotypes of each individual are listed in Supplementary Table S2.

A total of 11 haplotypes, Ht1–Ht11, could be inferred, and their relationships are depicted in Figure 1. Four additional haplotypes, r1–r4, resulted from four unique recombination events. Two of the recombinant haplotypes were observed in Africa, and one each in Europe and Southwest Asia. Ht1 was hypothesized to be the oldest haplotype in the phylogeny, since it was identical to the chimpanzee haplotype with the exception of *g.174351T>A* (GenBank L11910.1), which appeared to be recurrent in chimpanzee as all other primates carried the ancestral T allele. Arrows indicate the number and direction of sequence changes. Several intermediate haplotypes were not observed in the present study, because of either the relatively small size of the screening set or their disappearance due to random genetic drift or selection.

The color-coding in Figure 1 indicates the geographic distribution, and the areas of the circles are proportional to the frequencies of the haplotypes. Only one haplotype, Ht3, was



	Ht1	Ht2	Ht3	Ht4	Ht5	Ht6	Ht7	Ht8	Ht9	Ht10	Ht11	r1	r2	r3	r4	Total
Africa	5	5	5	1	0	0	0	4	3	2	5	3	0	0	1	34
West Asia	1	0	13	0	0	2	1	0	0	1	0	0	0	0	0	18
South Asia	0	0	48	0	1	14	1	0	1	3	2	0	2	0	0	72
Patient (S. Asia)	0	0	21	0	0	9	4	0	0	0	2	0	2	0	0	38
Europe	0	0	32	0	0	4	1	0	0	0	0	0	0	1	0	38
East Asia	0	0	25	0	0	10	0	0	0	0	3	0	0	0	0	38
America	0	0	14	0	0	5	0	0	0	1	0	0	0	0	0	20
Oceania	0	0	11	0	0	0	0	0	0	1	4	0	0	0	0	16
Total	6	5	169	1	1	44	7	4	4	8	16	3	4	1	1	274

FIGURE 1. **Phylogenetic relationship among *RB1* haplotypes. Primate sequences were used to determine the root of the phylogeny (Ht1). The circles each represent one haplotype, the area of the circle is proportional to the number of chromosomes observed with that haplotype, and the colors represent seven distinct geographic regions. The number of arrows indicates the number of changes required to generate the next related haplotype. The haplotypes were defined by the 13 congruent sites listed in Supplementary Table S2. There are 173,980 bp between the first and the last marker. Haplotypes are defined as follows (the base pair change defining a new haplotype is underlined): Ht1: CACGGGCTGAATT; Ht2: CACAGGCTGAATT; Ht3: CACAGGCTCAATT; Ht4: CATAGGCTCAATT; Ht5: AATAGGCTCAATT; Ht6: AATAGGCTCAATA; Ht7: ACTAGGCTCAATA; Ht8: CACAGACTGAGTT; Ht9: CACAGACTGAGCT; Ht10: CACATACGGAGCT; Ht11: CACGGGTTGGATT; r1: CACGGGTTGAGCT; r2: CACATACTCAATA; r3: CACAGGCTCAATA; and r4: CACGGGCTGAGTT. The table at the bottom of the figure lists the number of haplotypes observed in the different geographic regions.**

found in every region of the world, accounting for 61.7% of the haplotypes inferred. Its frequency was highest in Europe (84.2%), followed in decreasing order by West Asia (72.2%), America (70.0%), Oceania (68.8%), East Asia (65.8%), India (62.7%), and Africa, where it accounted only for 14.7% of all haplotypes observed. The second most frequent (16.1%) haplotype, Ht6, was found in every region of the world with the exception of Africa and Oceania. Its origin may have been in Southeast Asia, where Ht6 accounts for 26.3% (East Asia) to 20.9% (India) of the haplotypes inferred. Its frequency in the Americas is 25%, while it drops to 11.1% and 10.5%, respectively, in West Asia and Europe. The second most frequent (25%) haplotype in Oceania was Ht11. All other haplotypes outside Africa were observed at frequencies <8%. Like nucleotide diversity, haplotype diversity was greatest in Africa, where 8 out of 11 nonrecombinant haplotypes and two out of four recombinant haplotypes were found. Eight haplotypes were observed in India, five in West Asia, four in Europe, and three each in East Asia, Oceania, and America.

Haplotype 7 (Ht7), which contains *g. 1328A>C* (GenBank L11910.1) in the 5′UTR promoter region of *RB1*, was observed in 4 of the 19 retinoblastoma patients, but in none of the 25 Indian controls ($\chi^2=5.51$; $P=0.025$). Even upon inclusion of all individuals of Southwest Asian and European origin, a total of 64, in whom the derived C allele was observed three times (one Palestinian, one Baloochi, and one Georgian), the frequency of Ht7 in retinoblastoma patients remained significantly higher ($\chi^2=4.86$; $P=0.05$).

Table 4 summarizes the results of various quantitative measures of linkage disequilibrium obtained by pairwise comparisons between the 10 SNPs in the *RB1* region that had a minor allele frequency >5% in the total sample. Linkage disequilibrium is generally strong across the *RB1* region. Particularly in the non-African sample most values of $|D'|$ equaled 1. This indicates that SNP loci have not been separated by recombination, recurrent mutation, or gene conversion during the history of the sample [Ardlie et al. 2002]. This result was expected since all polymorphic sites could be accommodated by the same phylogenetic topology (Fig. 1). Although the non-African sample yielded more $|D'|$ values of 1 than the African sample, suggesting that linkage disequilibrium is stronger in the former, a formal statistical comparison of linkage disequilibrium between the two populations is not possible using this measure, because values of $|D'|$ less than 1 are strongly inflated in small samples containing SNPs with rare alleles. Another confounding factor in this study, which renders comparison of linkage disequilibria difficult, is the fact the African and non-African populations are composites of dozens of different ethnicities, most of which are represented by not more than three individuals. Consequently, differences in allele frequencies among the populations that are combined will lead to increases in linkage disequilibrium levels. The existence of such differences in allele frequency among populations is obvious from the Hardy-Weinberg (HW) monogenic disequilibrium coefficients M_{AA} and M_{BB} in Table 4 that were obtained by log-linear modeling using all individuals but the retinoblastoma patients. The polymorphism *g. 76721G>C* (GenBank L11910.1) in intron 14 showed a significant departure from HW equilibrium in the total sample of Africans and non-Africans combined. No significant departure from HW equilibrium, however, was observed when Africans and non-Africans were analyzed separately. There were also, with a single exception, no statistically significant differences in the monogenic disequilibrium coefficients τM_{AA} and τM_{BB} between the two populations (Table 4). The departure in HW equilibrium observed for *g. 76721G>C* in the total sample is most probably a result of genetic drift during the exodus of anatomically modern humans from Africa, in the process of which the derived C allele became the dominant allele in non-Africans, while the ancestral G remained the major allele in Africans. It is this locus that separates haplotypes Ht2 and Ht3 (Fig. 1).

Although r^2, another measure of linkage disequilibrium, shows much less inflation in a small sample than does $|D'|$, a low pairwise r^2 value is not necessarily indicative of high ancestral recombination between two loci, but instead may only reflect differences in allele frequencies at two loci [Ardlie et al., 2002]. For r^2 to equal 1, it is not only necessary for two markers not to have been separated by recombination, recurrent mutation, or gene conversion, but they also must have the same allele frequency. In the present study, this holds only true for three

TABLE 4. Pair-Wise Measures of Linkage Disequilibrium, Namely $|D'|$, r^2, S_{AB}, and Q_{AB}, and, Monogenic Disequilibrium Coefficients for 10 Selected SNP Markers from the *RBI* Region

	371	39606	41774	44591	65515	76721	153104	156616	173882	174351
$r^2\backslash\|D'\|$, Africans										
371		na	na	na	na	na	na	na	na	na
39606	na		na	na	na	na	na	na	na	na
41774	na	na		1.000	0.299	1.000	1.000	0.136	0.378	na
44591	na	na	0.252[a]		1.000	1.000	1.000	1.000	0.389	na
65515	na	na	0.763[b]	0.130		0.370	1.000	0.190	0.232	na
76721	na	na	0.015	0.077	0.011		1.000	1.000	1.000	na
153104	na	na	0.246[a]	0.062	0.479[b]	0.037		1.000	0.743	na
156616	na	na	0.008	0.582[b]	0.008	0.133	0.107		1.000	na
173882	na	na	0.063	0.129	0.046	0.066	0.029	0.497[b]		na
174351	na	na	na	na	na	na	na	na	na	
$r^2\backslash\|D'\|$, non-Africans										
371		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.975
39606	1.000[b]		1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.975
41774	0.015	0.015		1.000	1.000	1.000	1.000	1.000	1.000	1.000
44591	0.013	0.013	0.003		1.000	0.585	1.000	1.000	1.000	0.170
65515	0.013	0.013	0.913[b]	0.002		1.000	1.000	1.000	1.000	1.000
76721	0.024	0.024	0.612[b]	0.191[b]	0.559[b]		1.000	1.000	1.000	1.000
153104	0.013	0.013	0.913[b]	0.002	1.000[b]	0.559[b]		1.000	1.000	1.000
156616	0.008	0.008	0.002	0.625[b]	0.001	0.349[b]	0.001		1.000	1.000
173882	0.008	0.008	0.002	0.625[b]	0.001	0.349[b]	0.001	1.000[b]		1.000
174351	0.863[b]	0.863[b]	0.016	0.005	0.015	0.026	0.015	0.009	0.009	
$S_{AR}\backslash QAR$										
371		24.2[b]	3.31	2.22	2.14	0.38	1.14	0.22	1.68	3.43
39606	85.90[b]		3.31	2.39	2.00	0.04	1.14	3.13	1.54	3.43
41774	5.43	5.43		1.14	15.70[b]	38.60[b]	18.50[b]	4.34	4.59	4.65
44591	0.28	1.91	4.61		0.93	24.60[b]	0.58	21.90[b]	6.27	0.22
65515	8.62[a]	3.29	37.70[b]	0.10		21.90[b]	10.80[b]	2.47	7.61	3.01
76721	11.40[b]	5.43	26.90[b]	2.07	23.10[b]		16.90[b]	39.20[b]	22.00[b]	12.40[b]
153104	7.29	7.29	28.60[b]	0.16	46.50[b]	15.70[b]		0.79	0.26	1.87
156616	6.63	1.91	2.07	29.20[b]	0.10	8.07[a]	0.30		7.03	3.73
173882	5.99	1.50	2.66	23.50[b]	0.21	9.45[a]	0.47	57.00[b]		2.08
174351	71.50[b]	71.50[b]	10.30[a]	5.97	7.76[a]	4.54	6.56	1.50	1.15	
$M_{AA}, M_{BB}\backslash\tau M_{AA}, \tau M_{BB}$										
371		0.03	0.26	3.15	0.37	0.00	0.05	1.35	0.21	4.27
39606	0.68		0.26	2.58	0.30	9.02	0.05	0.22	0.15	4.27
41774	0.79	0.79		1.70	0.20	0.00	0.09	0.29	0.23	0.27
44591	3.00	3.00	3.11		3.25	0.25	2.94	0.07	0.95	0.07
65515	0.50	0.50	0.61	2.82		0.00	0.00	0.15	0.07	0.34
76721	20.20[b]	20.20[b]	15.50[b]	17.70[b]	15.20[b]		0.00	0.00	0.00	0.00
153104	1.16	1.16	1.27	3.48	0.98	15.80[b]		0.17	0.14	0.00
156616	5.55	5.55	5.66	7.87[a]	5.37	2020[b]	6.02		0.08	0.15
173882	1.17	1.17	1.28	3.49	0.99	15.80[b]	0.65	6.04		
174351	0.43	0.43	3.35	5.56	3.06	17.90[b]	3.72	8.10[a]	3.73	

[a]$P < 0.05$.
[b]$P < 0.01$.

pairs of markers in the non-African population, namely, *g. 317C>A* and *g. 39606C>T*, *g. 65515C>T* and *g. 153104A>G*, and *g. 156616A>G* and *g. 173882T>C*.

In an attempt to test for statistical significance of differences in levels of linkage disequilibria between Africans and non-Africans, we used log-linear modeling [Huttley and Wilson, 2000]. The terms S_{AB} and Q_{AB} are the sum and the product, respectively, of digenic disequilibria for the total sample. Log-linear modeling shows that 33 to 45% of all possible bilocus combinations depart significantly from linkage equilibrium (Table 4). However, due to the small number of Africans and the strong linkage disequilibria among loci in the *RB1* region, there was insufficient power to test for differences in digenic disequilibria between the two populations.

DISCUSSION

In this study, apparently pathogenic mutations were detected in 9 of 19 (48%) patients with hereditary retinoblastoma. Seven of the nine mutations were single-base substitutions, one was a small deletion, and the remaining one was a complex mutation. Six of the nine mutations had not been reported previously (www.d-lohmann.de/Rb/mutations.html). The functional significance of a 2-bp deletion in intron 3 and a T>A transversion in intron 6, which have been detected only once each in two patients in whom no other mutations had been identified, remains to be elucidated. Although the high mutation detection sensitivity of DHPLC ($\geq$95%) has been documented in numerous studies [Xiao and Oefner 2001], we decided to sequence all 19 retinoblastoma cases to ensure that mutations had not been missed by DHPLC. Indeed, no additional mutations were detected by dye-terminator sequencing.

The percentage of cases, in which a pathogenic mutation could be identified in the present study (47%), compares well to other studies that also did not apply methodology to detect large genomic rearrangements. Blanquet et al. [1993, 1994, 1995] detected mutations in 25.5% of 176 hereditary retinoblastoma patients, while Cowell et al. [1994] and Liu et al. [1995] found

mutations in 48% of their patients with hereditary retinoblastoma. Recent studies from India reported in unrelated samples mutations in 7 out of 21 (33%) [Ata-ur-Rasheed et al., 2002] and 22 out of 47 (47%) [Kiran et al., 2003] using sequencing and SSCP of PCR-amplified exons, respectively. That large hemizygous deletions are responsible for the majority of the remaining cases of hereditary retinoblastoma has been confirmed in studies employing, in addition to mutational screening by SSCP, heteroduplex analysis, and direct sequencing, either by Southern blot hybridization [Kloss et al., 1991; Lohmann et al., 1992, 1994a, 1994b, 1996] or quantitative multiplex PCR [Houdayer et al., 2004; Richter et al., 2003]. Kloss et al. [1991] and Lohmann et al. [1992, 1994a, 1994b, 1996] identified mutations in 99 out of 119 (83%) patients with hereditary retinoblastoma, while Richter et al. [2003] detected mutations in 89% of bilaterally affected probands and both mutant alleles in 84% of tumors from unilaterally affected probands. Finally, Houdayer et al. [2004] detected *RB1* gene mutations in 81.5% (83/102) of the bilateral and/or familial cases and 5.5% (5/90) of unilateral sporadic retinoblastoma patients. Therefore, we conclude that the remaining cases of bilateral retinoblastoma, for which we have failed to detect an apparently disease causing mutation by DHPLC and sequencing, are most probably the result of large deletions. Alternatively, hypermethylation of the CpG island within the promoter region [Greger et al., 1994; Stirzaker et al., 1997], mosaicism [Sippel et al., 1998], or mutations within intronic sequences not covered by the present study, cannot be excluded either as a cause of retinoblastoma.

Including 5&UTR and introns, 46% (N=23) of the apparently nonpathogenic polymorphisms were observed only once and 17 were located in the noncoding region. The frequency of singletons is lower than in *ATM* [Thorstenson et al., 2001], *BRCA2* [Wagner et al., 1999], and *CSTB* [Sale et al., 2002], in which they constituted 60, 59, and 67% of all polymorphic sites, respectively. However, it is higher than the frequency of singletons detected in an identical set of individuals in *WRN* [Passarino et al., 2001] and *TUBB* [Sale et al., 2002], in which only 33 and 22%, respectively, of all variant sites were observed once. Only seven (14%) of the *RB1* sequence variants had a heterozygosity $\geq$10%, which is similar to *BRCA2* [Wagner et al., 1999] and *ATM* [Thorstenson et al., 2001], with 16% of their variants being heterozygous in $\geq$10% of the individuals screened. It is, however, much less than that in the lipoprotein lipase gene [Nickerson et al., 1998].

Overall, sequence diversity was three times lower in non-Africans than Africans. Generally, it is assumed that the lower sequence diversity in non-African populations is the consequence of bottlenecks that occurred during the exodus from Africa and the colonization of continents. However, data on the loss rate of mitochondrial and nuclear alleles in the Saguenay population in northern Quebec suggest that in an expanding population, low sequence diversity can occur even without any bottleneck phenomenon [Heyer, 1995]. Particularly in the case of mitochondrial DNA, this was explained by near neutrality [Nachman et al., 1996; Rand and Kann, 1996]. Under the strictly neutral theory, it is expected that heterozygosity increases with effective population size [Kimura, 1968]. In the near neutral theory, however, which has been invoked to account for excess amino acid variation in the mitochondrial DNA of humans and mice, selection becomes more effective the larger the population and, consequently, heterozygosity decreases [Nei, 1987]. Interestingly, this scenario may also apply to *RB1* for which the ratio of replacement to silent nucleotide substitutions is higher, though not statistically significant, within human than between species. This is inconsistent with a strictly neutral model and suggests that the amino acid variants observed in human *RB1* are slightly deleterious. Under this scenario, polymorphism is governed by selection coefficients that are neither sufficiently strong such that replacement mutations would be eliminated rapidly from the population, nor sufficiently weak such that they could persist in the population long enough to become fixed and contribute to sequence divergence between species. The low frequency of the derived *RB1* alleles further supports this scenario. An alternative explanation for the data is that there has been a recent relaxation of selective constraint in the human lineage. Given the function of Rb protein in cell cycle regulation [Harbour, 2001], the latter explanation appears less likely.

The most striking observation about *RB1* in the present study is the 50-fold smaller nucleotide diversity in the coding compared to the noncoding region. This is unique among the 20 genes studied in the same set of individuals. In accordance with other studies [Cargill et al., 1999; Glatt et al., 2001; Halushka et al., 1999], ratios of nucleotide diversity estimates for coding and noncoding regions of the majority of genes range from 1:1 to 1:2.5. To date, only *ATM* [Thorstenson et al., 2001] and *TUBB* [Sale, et al., 2002] had shown greater ratios, 1:7.5 and 1:6.3, respectively. Moreover, the nucleotide diversity estimate for the coding region of *RB1* is the smallest reported for an autosomal gene. This suggested strong pressure to keep the amino acid sequence of *RB1* conserved. To further support this hypothesis, we sequenced the exons and flanking intronic sequences of *RB1* in five chimpanzees, two lowland gorillas, two orangutans, one rhesus monkey, and one spider monkey. In all primates except chimpanzee fixed silent differences were more frequent than amino acid replacement substitutions in comparison to human, which is generally taken as evidence of purifying selection. Although the HKA test did not reach the level of significance in any of the comparisons, the hypothesis that stronger purifying selection has been acting on *RB1* is supported by the low ratio of nonsynonymous to synonymous mutations of 0.095 relative to other autosomal genes such as ATM, in which the ratio is estimated to be 0.18. Given the key role of *RB1* in cell cycle control, this high degree of sequence conservation across species does not come as a surprise. *RB1* is the third gene in our hands aside from *ATM* [Thorstenson et al., 2001] and *TUBB* [Sale et al., 2002] that appears to be under the influence of purifying selection. However, what sets *RB1* apart from the latter is the paucity of synonymous polymorphisms not only in human but also in primates. This suggests a significant bias toward a subset of preferred codons. A significant correlation between translational efficiency and codon usage has been found in a number of organisms, including *Escherichia coli*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Rattus norvegicus*, and *Mus musculus* [Konu and Li, 2002]. Somatic mutations in *RB1*, resulting in its inactivation and consequent loss of proliferation control, are a common event in many cancers [Sherr, 1996]. Although it has been suggested that partial inactivation of Rb may constitute the molecular basis of low-penetrance retinoblastoma [Harbour, 2001], no studies have been conducted to test the hypothesis whether germline synonymous polymorphisms predispose to cancer by reducing its expression. The present population genetic data seems to render such a hypothesis worth testing.

The prevalence of a single haplotype outside Africa, accounting for more than two-thirds of all haplotypes observed, suggests that,

in the absence of a plausible reason for the existence of a greater selection coefficient in non-Africans, this is the result of genetic drift during the dispersal of anatomically modern humans from Africa, which was accompanied by a severe bottleneck with an effective size estimated previously at a few hundred individuals over a period of up to 400 generations [Reich et al., 2001].This is not the first time that we have observed a significant reduction in haplotype diversity. In *TUBB*, a single haplotype accounted globally for 88% of all haplotypes [Sale et al., 2002]. In *ATM*, on the other hand, two haplotypes characterized in fairly equal proportion 84% of the sample [Thorstenson, et al., 2001], while three haplotypes in a 565-bp chromosome 21 region represented 23, 29, and 30% of 708 chromosomes analyzed worldwide [Jin, et al., 1999]. Recent analysis of haplotype diversity over the entire human chromosome 21 [Patil et al., 2001] has revealed a significant number of blocks in which more than 80% of a global human sample can typically be characterized by only three common haplotypes. However, what sets *RB1* aside from other genes is the distance over which linkage disequilibrium extends in both Africans and non-Africans (Table 4). This is most likely a consequence of directional selection acting on *RB1*. Among autosomal genes, a similar extent of linkage disequilibrium has been only reported for the *WASL* (MIM# 605056) gene [Reich et al., 2001] and *ATM* [Thorstenson et al., 2001]. Although it was not possible to test formally for a difference in the levels of linkage disequilibria between Africans and non-Africans, the smaller fraction of $|D'|$ values of 1 (Table 4) in the former suggests that linkage disequilibrium decays at a faster rate in African than in non-African populations, which concords with other studies [Hollox et al., 2001; Kidd et al., 1998, 2000; Mateu et al., 2001; Reich et al., 2001; Tishkoff et al., 1996, 1998].

ACKNOWLEDGMENTS

We thank the DNA donors and the investigators who provided the samples of human genomic DNA: L.L. Cavalli-Sforza, L. Excoffier, M.E. Ibrahim, T. Jenkins, J. Kidd, A. Langaney, S.Q. Mehdi, P. Parham, and R. Rajalingam. We thank Marc Feldman, Charles Langley, and Eva Lee for helpful discussions, as well as Gavin Huttley for his assistance in implementing log-linear models as a test for different levels of linkage disequilibria between populations. P.J.O. holds U.S. patents related to DHPLC and receives annual royalty payments from Transgenomic and Varian.

REFERENCES

Alonso J, Garcia-Miguel P, Abelairas J, Mendiola M, Sarret E, Vendrell MT, Navajas A, Pestana A. 2001. Spectrum of germline RB1 gene mutations in Spanish retinoblastoma patients: phenotypic and molecular epidemiological implications. Hum Mutat 17:412–422.

Altshuler D, Pollara VJ, Cowles CR, Van Etten WJ, Baldwin J, Linton L, Lander ES. 2000. An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature 407:513–516.

Ardlie KG, Kruglyak L, Seielstad M. 2002. Patterns of linkage disequilibrium in the human genome. Nat Rev Genet 3: 299–309.

Ata-ur-Rasheed M, Vemuganti G, Honavar S, Ahmed N, Hasnain S, Kannabiran C. 2002. Mutational analysis of the RB1 gene in Indian patients with retinoblastoma. Ophthalmic Genet 23:121–128.

Blanquet V, Turleau C, Gross MS, Goossens M, Besmond C. 1993. Identification of germline mutations in the RB1 gene by denaturant gradient gel electrophoresis and polymerase chain reaction direct sequencing. Hum Mol Genet 2:975–979.

Blanquet V, Gross MS, Turleau C, Senamaud-Beaufort C, Doz F, Besmond C. 1994. Three novel germline mutations in exons 8 and 18 of the retinoblastoma gene. Hum Mol Genet 3: 1185–1186.

Blanquet V, Turleau C, Gross-Morand MS, Senamaud-Beaufort C, Doz F, Besmond C. 1995. Spectrum of germline mutations in the RB1 gene: a study of 232 patients with hereditary and non hereditary retinoblastoma. Hum Mol Genet 4:383–388.

Bulmer M. 1991. The selection-mutation-drift theory of synonymous codon usage. Genetics 129:897–907.

Cambien F, Poirier O, Nicaud V, Herrmann SM, Mallet C, Ricard S, Behague I, Hallet V, Blanc H, Loukaci V, Thillet J, Evans A, Ruidavets JB, Arveiler D, Luc G, Tiret L. 1999. Sequence diversity in 36 candidate genes for cardiovascular disorders. Am J Hum Genet 65:183–191.

Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Shaw N, Lane CR, Lim EP, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley GQ, Lander ES. 1999. Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nat Genet 22:231–238.

Choy KW, Pang CP, Yu CB, Wong HL, Ng JS, Fan DS, Lo KW, Chai JT, Wang J, Fu W, Lam DS. 2002. Loss of heterozygosity and mutations are the major mechanisms of RB1 gene inactivation in Chinese with sporadic retinoblastoma. Hum Mutat 20:408.

Collins FS, Guyer MS, Charkravarti A. 1997. Variations on a theme: cataloging human DNA sequence variation. Science 278:1580–1581.

Cowell JK, Smith T, Bia B. 1994. Frequent constitutional C to T mutations in CGA-arginine codons in the RB1 gene produce premature stop codons in patients with bilateral (hereditary) retinoblastoma. Eur J Hum Genet 2:281–290.

Cowell JK, Cragg H. 1996. Constitutional nonsense germline mutations in the RB1 gene detected in patients with early onset unilateral retinoblastoma. Eur J Cancer 32A:1749–1752.

Dalamon V, Surace E, Borelina D, Ziembar M, Esperante S, Francipane L, Davila M, Parma D, Szijan I. 2001. Detection of mutations in argentine retinoblastoma patients by segregation of polymorphisms, exon analysis and cytogenetic test. Ophthalmic Res 33:336–339.

Dunn JM, Phillips RA, Zhu X, Becker A, Gallie BL. 1989. Mutations in the RB1 gene and their effects on transcription. Mol Cell Biol 9:4596–4604.

Fung YK, Murphree AL, T'Ang A, Qian J, Hinrichs SH, Benedict WF. 1987. Structural evidence for the authenticity of the human retinoblastoma gene. Science 236:1657–1661.

Glatt CE, DeYoung JA, Delgado S, Service SK, Giacomini KM, Edwards RH, Risch N, Freimer NB. 2001. Screening a large reference sample to identify very low frequency sequence variants: comparisons between two genes. Nat Genet 27: 435–438.

Goddard AD, Balakier H, Canton M, Dunn J, Squire J, Reyes E, Becker A, Phillips RA, Gallie BL. 1988. Infrequent genomic rearrangement and normal expression of the putative RB1 gene in retinoblastoma tumors. Mol Cell Biol 8:2082–2088.

Greger V, Debus N, Lohmann D, Hopping W, Passarge E, Horsthemke B. 1994. Frequency and parental origin of hypermethylated RB1 alleles in retinoblastoma. Hum Genet 94:491–496.

Halushka MK, Fan JB, Bentley K, Hsie L, Shen N, Weder A, Cooper R, Lipshutz R, Chakravarti A. 1999. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. Nat Genet 22:239–247.

Harbour JW. 2001. Molecular basis of low-penetrance retinoblastoma. Arch Ophthalmol 119:1699–1704.

Harris EE, Hey J. 1999. X chromosome evidence for ancient human histories. Proc Natl Acad Sci USA 96: 3320–3324.

Hartl DL, Clark AG. 1997. Principles of population genetics. Sunderland: Sinauer Associates, p 95–107.

Hasegawa M, Cao Y, Yang Z. 1998. Preponderance of slightly deleterious polymorphism in mitochondrial DNA: nonsynonymous/synonymous rate ratio is much higher within species than between species. Mol Biol Evol 15:1499–1505.

Heyer E. 1995. Mitochondrial and nuclear genetic contribution of female founders to a contemporary population in northeast Quebec. Am J Hum Genet 56:1450–1455.

Hollox EJ, Poulter M, Zvarik M, Ferak V, Krause A, Jenkins T, Saha N, Kozlov AI, Swallow DM. 2001. Lactase haplotype diversity in the Old World. Am J Hum Genet 68: 160–172.

Houdayer C, Gauthier-Villars M, Lauge A, Pages-Berhouet S, Dehainault C, Caux-Moncoutier V, Karczynski P, Tosi M, Doz F, Desjardins L, Couturier J, Stoppa-Lyonnet D. 2004. Comprehensive screening for constitutional RB1 mutations by DHPLC and QMPSF. Hum Mutat 23:193–202.

Hudson RR, Kreitman M, Aguade M. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics 116:153–159.

Huttley GA, Wilson SR. 2000 Testing for concordant equilibria between population samples. Genetics 156:2127–2135.

Jakubowska A, Zajaczek S, Haus O, Limon J, Kostyk E, Krzystolik Z, Lubinski J. 2001. Novel RB1 gene constitutional mutations found in Polish patients with familial and/or bilateral retinoblastoma. Hum Mutat 18:459.

Jin L, Underhill PA, Doctor V, Davis RW, Shen P, Cavalli-Sforza LL, Oefner PJ. 1999. Distribution of haplotypes from a chromosome 21 region distinguishes multiple prehistoric human migrations. Proc Natl Acad Sci USA 96:3796–3800.

Kidd KK, Morar B, Castiglione CM, Zhao H, Pakstis AJ, Speed WC, Bonne-Tamir B, Lu RB, Goldman D, Lee C, Nam YS, Grandy DK, Jenkins T, Kidd JR. 1998. A global survey of haplotype frequencies and linkage disequilibrium at the DRD2 locus. Hum Genet 103:211–227.

Kidd JR, Pakstis AJ, Zhao H, Lu RB, Okonofua FE, Odunsi A, Grigorenko E, Tamir BB, Friedlaender J, Schulz LO, Parnas J, Kidd KK. 2000. Haplotypes and linkage disequilibrium at the phenylalanine hydroxylase locus, PAH, in a global representation of populations. Am J Hum Genet 66:1882–1899.

Kimura M. 1968. Evolutionary rate at the molecular level. Nature 217:624–626.

Kiran VS, Kannabiran C, Chakravarthi K, Vemuganti GK, Honavar SG. 2003. Mutational screening of the RB1 gene in Indian patients with retinoblastoma reveals eight novel and several recurrent mutations. Hum Mutat 22:339.

Kloss K, Wahrisch P, Greger V, Messmer E, Fritze H, Hopping W, Passarge E, Horsthemke B. 1991. Characterization of deletions at the retinoblastoma locus in patients with bilateral retinoblastoma. Am J Med Genet 39:196–200.

Klutz M, Horsthemke B, Lohmann DR. 1999. RB1 gene mutations in peripheral blood DNA of patients with isolated unilateral retinoblastoma. Am J Hum Genet 64: 667–668.

Knudson AG, Jr. 1971. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci USA 68:820–823.

Konu O, Li MD. 2002. Correlations between mRNA expression levels and GC contents of coding and untranslated regions of genes in rodents. J Mol Evol 54:35–41.

Kruglyak L. 1997. The use of a genetic map of biallelic markers in linkage studies. Nat Genet 17:21–24.

Kumaramanickavel G, Joseph B, Narayana K, Natesh S, Mamatha G, Shanmugam MP, Elamparathi A, Biswas J. 2003. Molecular-genetic analysis of two cases with retinoblastoma: benefits for disease management. J Genet 82:39–44.

Lander ES. 1996. The new genomics: global views of biology. Science 274:536–539.

Lee WH, Bookstein R, Hong F, Young LJ, Shew JY, Lee EY. 1987a. Human retinoblastoma susceptibility gene: cloning, identification, and sequence. Science 235:1394–1399.

Lee WH, Shew JY, Hong FD, Sery TW, Donoso LA, Young LJ, Bookstein R, Lee EY. 1987b. The retinoblastoma susceptibility gene encodes a nuclear phosphoprotein associated with DNA binding activity. Nature 329:642–645.

Liu Z, Song Y, Bia B, Cowell JK. 1995. Germline mutations in the *RB1* gene in patients with hereditary retinoblastoma. Genes Chromosomes Cancer 14:277–284.

Lohmann D, Horsthemke B, Gillessen-Kaesbach G, Stefani FH, Hofler H. 1992. Detection of small *RB1* gene deletions in retinoblastoma by multiplex PCR and high-resolution gel electrophoresis. Hum Genet 89:49–53.

Lohmann DR, Brandt B, Hopping W, Passarge E, Horsthemke B. 1994a. Distinct RB1 gene mutations with low penetrance in hereditary retinoblastoma. Hum Genet 94: 349–354.

Lohmann DR, Brandt B, Hopping W, Passarge E, Horsthemke B. 1994b. Spectrum of small length germline mutations in the *RB1* gene. Hum Mol Genet 3:2187–2193.

Lohmann DR, Brandt B, Hopping W, Passarge E, Horsthemke B. 1996. The spectrum of RB1 germ-line mutations in hereditary retinoblastoma. Am J Hum Genet 58:940–949.

Lohmann DR, Gerick M, Brandt B, Oelschlager U, Lorenz B, Passarge E, Horsthemke B. 1997. Constitutional RB1-gene mutations in patients with isolated unilateral retinoblastoma. Am J Hum Genet 61:282–294.

Lohmann DR. 1999. RB1 gene mutations in retinoblastoma. Hum Mutat 14:283–288.

Mateu E, Sanchez F, Najera C, Beneyto M, Castell V, Hernandez M, Serra I, Prieto F. 1997. Genetics of retinoblastoma: a study. Cancer Genet Cytogenet 95:40–50.

Mateu E, Calafell F, Lao O, Bonne-Tamir B, Kidd JR, Pakstis AJ, Kidd KK, Bertranpetit J. 2001. Worldwide genetic analysis of the *CFTR* region. Am J Hum Genet 68:103–117.

Nachman MW, Brown WM, Stoneking M, Aquadro CF. 1996. Nonneutral mitochondrial DNA variation in humans and chimpanzees. Genetics 142:953–963.

Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.

Nickerson DA, Taylor SL, Weiss KM, Clark AG, Hutchinson RG, Stengard J, Salomaa V, Vartiainen E, Boerwinkle E, Sing CF. 1998. DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. Nat Genet 19:233–240.

Nielsen R, Yang Z. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148:929–936.

Passarino G, Shen P, Van Kirk JB, Lin AA, De Benedictis G, Cavalli Sforza LL, Oefner PJ, Underhill PA. 2001. The Werner syndrome gene and global sequence variation. Genomics 71:118–122.

Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, Kautzer CR, Lee DH, Marjoribanks C, McDonough DP, Nguyen BT, Norris MC, Sheehan JB, Shen N, Stern D, Stokowski RP, Thomas DJ, Trulson MO, Vyas KR, Frazer KA, Fodor SP, Cox DR. 2001. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science 294:1719–1723.

Plotkin JB, Robins H, Levine AJ. 2004. Tissue-specific codon usage and the expression of human genes. Proc Natl Acad Sci USA 101:12588–12591.

Rand DM, Kann LM. 1996. Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from Drosophila, mice, and humans. Mol Biol Evol 13:735–748.

Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES. 2001. Linkage disequilibrium in the human genome. Nature 411:199–204.

Richter S, Vandezande K, Chen N, Zhang K, Sutherland J, Anderson J, Han L, Panton R, Branco P, Gallie B. 2003. Sensitive and efficient detection of RB1 gene mutations enhances care for families with retinoblastoma. Am J Hum Genet 72:253–269.

Rieder MJ, Taylor SL, Clark AG, Nickerson DA. 1999. Sequence variation in the human angiotensin converting enzyme. Nat Genet 22:59–62.

Sale S, Sung R, Shen P, Yu K, Wang Y, Duran GE, Kim JH, Fojo T, Oefner PJ, Sikic BI. 2002. Conservation of the class I beta-tubulin gene in human populations and lack of mutations in lung cancers and paclitaxel-resistant ovarian cancers. Mol Cancer Ther 1:215–225.

Schubert EL, Hansen MF. 1996. A previously unknown polymorphism located within the RB1 locus only present in Asian individuals. Hum Hered 46:118–120.

Shen P, Wang F, Underhill PA, Franco C, Yang WH, Roxas A, Sung R, Lin AA, Hyman RW, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ. 2000. Population genetic implications from sequence variation in four Y chromosome genes. Proc Natl Acad Sci USA 97:7354–7359.

Shen P, Buchholz M, Sung R, Roxas A, Franco C, Yang WH, Jagadeesan R, Davis K, Oefner PJ. 2002. Population genetic implications from DNA polymorphism in random human genomic sequences. Hum Mutat 20:209–217.

Sherr CJ. 1996. Cancer cell cycles. Science 274:1672–1677.

Sippel KC, Fraioli RE, Smith GD, Schalkoff ME, Sutherland J, Gallie BL, Dryja TP. 1998. Frequency of somatic and germ-line mosaicism in retinoblastoma: implications for genetic counseling. Am J Hum Genet 62:610–619.

Stirzaker C, Millar DS, Paul CL, Warnecke PM, Harrison J, Vincent PC, Frommer M, Clark SJ. 1997. Extensive DNA methylation spanning the Rb promoter in retinoblastoma tumors. Cancer Res 57:2229–2237.

Thorstenson YR, Shen P, Tusher VG, Wayne TL, Davis RW, Chu G, Oefner PJ. 2001. Global analysis of ATM polymorphism reveals significant functional constraint. Am J Hum Genet 69:396–412.

Tiret L, Poirier O, Nicaud V, Barbaux S, Herrmann SM, Perret C, Raoux S, Francomme C, Lebard G, Tregouet D, Cambien F. 2002. Heterogeneity of linkage disequilibrium in human genes has implications for association studies of common diseases. Hum Mol Genet 11:419–429.

Tishkoff SA, Dietzsch E, Speed W, Pakstis AJ, Kidd JR, Cheung K, Bonne-Tamir B, Santachiara-Benerecetti AS, Moral P, Krings M. 1996. Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. Science 271: 1380–1387.

Tishkoff SA, Goldman A, Calafell F, Speed WC, Deinard AS, Bonne-Tamir B, Kidd JR, Pakstis AJ, Jenkins T, Kidd KK. 1998. A global haplotype analysis of the myotonic dystrophy locus: implications for the evolution of modern humans and for the origin of myotonic dystrophy mutations. Am J Hum Genet 62:1389–1402.

Toguchida J, McGee TL, Paterson JC, Eagle JR, Tucker S, Yandell DW, Dryja TP. 1993. Complete genomic sequence of the human retinoblastoma susceptibility gene. Genomics 17: 535–543.

Twells RC, Mein CA, Phillips MS, Hess JF, Veijola R, Gilbey M, Bright M, Metzker M, Lie BA, Kingsnorth A, Gregory E, Nakagawa Y, Snook H, Wang WY, Masters J, Johnson G, Eaves I, Howson JM, Clayton D, Cordell HJ, Nutland S, Rance H, Carr P, Todd JA. 2003. Haplotype structure, LD blocks, and uneven recombination within the LRP5 gene. Genome Res 13:845–855.

Urrutia AO, Hurst LD. 2003. The signature of selection mediated by expression on human genes. Genome Res 13:2260–2264.

Wagner TM, Hirtenlehner K, Shen P, Moeslinger R, Muhr D, Fleischmann E, Concin H, Doeller W, Haid A, Lang AH, Mayer P, Petru E, Ropp E, Langbauer G, Kubista E, Scheiner O, Underhill P, Mountain J, Stierer M, Zielinski C, Oefner P. 1999. Global sequence diversity of BRCA2: analysis of 71 breast cancer families and 95 control individuals of worldwide populations. Hum Mol Genet 8: 413–423.

Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E, Robinson E, Mittmann M, Morris MS, Shen N, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ, Lander ES. 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280:1077–1082.

Xiao W, Oefner PJ. 2001. Denaturing high-performance liquid chromatography: a review. Hum Mutat 17:439–474.

Yandell DW, Dryja TP. 1989. Detection of DNA sequence polymorphisms by enzymatic amplification and direct genomic sequencing. Am J Hum Genet 45:547–555.

Yandell DW, Campbell TA, Dayton SH, Petersen R, Walton D, Little JB, McConkie-Rosell A, Buckley EG, Dryja TP. 1989. Oncogenic point mutations in the human retinoblastoma gene: their application to genetic counseling. N Engl J Med 321: 1689–1695.

Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol 15:568–573.

Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol Evol 19:908–917.

Yilmaz S, Horsthemke B, Lohmann DR. 1998. Twelve novel RB1 gene mutations in patients with hereditary retinoblastoma. Mutation in Brief #206. Online. Hum Mutat 12:434.

Zhang Q, Minoda K, Zeng R, Wu Z, Xiao X, Li S, Zhang F. 1997. Exon-by-exon screening for RB germline mutations using Heteroduplex-SSCP analysis. Yan Ke Xue Bao 13:5–11.