Positive Selection on the Gene RNASEL: Correlation between Patterns of Evolution and Function

Wei Jin,†1 Dong-Dong Wu,*†2 Xin Zhang,1 David M. Irwin,2,3,4 and Ya-Ping Zhang*1,2

1Laboratory for Conservation and Utilization of Bio-resource, Yunnan University, Kunming, China
2State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China
3Department of Laboratory Medicine and Pathobiology, University of Toronto, Ontario, Canada
4Banting and Best Diabetes Centre, University of Toronto, Ontario, Canada
†These authors contributed equally to this work.
*Corresponding author: E-mail: dongdongw86@yahoo.com.cn; zhangyp@mail.kiz.ac.cn.
Associate editor: Willie Swanson

Abstract

RNASEL is a 2-5A-dependent endoribonuclease that is a component of the interferon-induced 2-5A system, which plays a crucial role in the antiviral and apoptotic activities of interferons. In humans, many polymorphic sites within the RNASEL gene have been associated with an increased risk of developing prostate cancer. Here, we obtained coding sequences for the RNASEL gene from 11 primates and found evidence that positive selection has operated on the C-terminal endoribonuclease domain and the N-terminal ankyrin repeats domain of the protein, domains that directly interact with virus (i.e., ankyrin repeats are responsible for receiving environmental signals, and the endoribonuclease catalyses the destruction of the pathogenic viral RNA). To extend this finding, we studied variation within this gene in modern human populations by resequencing alleles from 144 individuals representing four separate populations. Interestingly, the frequency of the 541D allele shows a negative association with the incidence rate of prostate cancer in worldwide populations, and haplotypes containing the 541D polymorphisms demonstrate signatures of positive selection. RNASEL variants having the 541D haplotype likely have a greater ability to defend against infections by viruses, thus the loss of this activity may be associated with the development of prostate cancer. We provide evidence that positive selection has operated on the RNASEL gene, and its evolution is correlated with its function in pathogen defense and cancer association.

Key words: RNASEL, positive selection, prostate cancer.

Introduction

RNASEL is a 2-5A-dependent endoribonuclease that is part of the interferon-induced 2-5A system, which plays a crucial role in the antiviral and apoptotic activities of interferons. When cells are infected by viruses, interferons induce the expression of OAS (2′, 5′-oligoadenylate synthetase), which will synthesize 2-5A (5′-triphosphorylated 2′, 5′-oligoadenylates) to activate RNASEL, an enzyme that degrades viral RNA (Hassel et al. 1993; Bisbal et al. 1995). The role of RNASEL has been demonstrated by RNASEL−/− mice that have an enhanced susceptibility to infections by the picornaviruses encephalomyocarditis virus and Coxsackievirus B4 (Zhou et al. 1997; Flodstrom-Tullberg et al. 2005).

The RNASEL gene has been found only in reptiles, birds, and mammals (Cayley et al. 1982; Silverman 2003). The RNASEL protein contains three different domains, an N-terminal nine ankyrin repeat domain, a protein kinase–like domain, and an endoribonuclease domain at the C-terminus. Comparison of the mouse and human RNASEL genes supports the idea that RNASEL evolved from the related stress-response protein Ire1s by recombination with an ankyrin repeat protein (Zhou et al. 2000).

RNASEL gene is also a candidate tumor suppressor gene as it maps to the hereditary prostate cancer locus at 1q24-q25 in the human genome (Smith et al. 1996). Many mutations, such as R462Q, D541E, and E265X located in the kinase region of human RNASEL, have been identified as increasing the risk of developing prostate cancer by affecting the enzymatic activity of the protein (Carpten et al. 2002; Casey et al. 2002; Rokman et al. 2002). Homozygosity for a derived allele, for example, Q with single nucleotide polymorphism (SNP) R462Q, reduces the activity of RNASEL by as much as 3-fold due to an impairment in enzyme dimerization (Casey et al. 2002). Men heterozygous for the Q/R alleles have a 50% increased risk of prostate cancer compared with the R/R genotype, and Q/Q homozygotes have more than double the risk (Casey et al. 2002). A meta-analysis of available data concluded that the 541E allele increased the risk of prostate cancer for Caucasian men, regardless of a family history of the disease (Li and Tai 2006). In addition, in African Americans, the 541D allele is associated with a decrease in the risk for sporadic prostate cancer. In particular, the 462R–541D haplotype shows a strong protective effect against prostate cancer (Robbins et al. 2008).
Many genes involved in the immune system have been demonstrated to have undergone positive selection (Vallender and Lahn 2004; Sabeti et al. 2006). Adaptive evolution of these genes can be explained by Van Valen’s (1973) Red Queen hypothesis where an “arms race” occurs between the host and pathogens. Since RNASEL is also involved in immunity, there is potential that it also evolves rapidly; indeed, RNASEL has been reported, based on a few sequences, to be under positive selection (Bustamante et al. 2005; Summers and Crespi 2008). To better study the evolution of RNASEL, we sequenced the RNASEL coding sequence from several primate species and used these sequences for maximum likelihood (ML)–based analyses. We also studied the variation in RNASEL sequences in modern humans by sequencing alleles from 144 individuals from four separate populations.

### Materials and Methods

**Samples, PCR, and Sequencing**
The cDNA sequences RNASEL from human (Homo sapiens), chimpanzee (Pan troglodytes), Sumatran orangutan (Pongo abelii), rhesus macaque (Macaca mulatta), horse (Equus caballus), cattle (Bos taurus), pig (Sus scrofa), dog (Canis lupus familiaris), mouse (Mus musculus), and rat (Rattus norvegicus) were obtained from NCBI. Wise2 (Binney et al. 2004) was used to predict the cDNA sequence of RNASEL from the marmoset (Callithrix jacchus) genome using gene sequences obtained by a BLAT search of the University of California–San Cruz (UCSC) genome data (http://genome.ucsc.edu/). Coding sequences of the RNASEL gene were obtained from six additional primate species, including the Tibetan macaque (Macaca thibetana), douc langur (Pygathrix nemaeus), long-tailed macaque (Macaca fascicularis), hoolock whitebrowed gibbon (Bunopithecus hoolock), black snub-nosed monkey (Pygathrix nemaeus), and olive baboon (Papio anubis), by polymerase chain reaction (PCR) and DNA sequencing. The mammalian sequences are summarized in supplementary material 1 (Supplementary Material online). In addition, 144 unrelated human individuals were chosen at random from the Human Genome Diversity Cell Line Panel (Cann et al. 2002) representing four human populations: 34 Africans, 37 Europeans, 37 East Asians, and 36 South Asians and sequenced for four regions covering the coding sequences (total 4,181 bp) of the RNASEL gene (supplementary fig. 1, Supplementary Material online). Human sequence data are summarized in supplementary material 2 (Supplementary Material online).

DNA sequencing was performed on an ABI 3730 automated DNA sequencer. Sequences were analyzed by DNASTar software. Sequences from two individuals containing deletions in region A were sequenced after cloning. PCR primers, and conditions and primers for sequencing are available upon request.

**Genotypes of SNPs R462Q, E541D, and Association with Age-Standardized Incidence Rate of Prostate Cancer**

Genotypes of SNP rs627928 (E541D) in unrelated individuals from the worldwide panel of the Human Genome Diversity Cell Line Panel (Cann et al. 2002) were obtained by independent PCR, cut by the restriction enzyme PsuI, and the products visualized after separation on an agarose gel. Genotypes of SNP rs486907 (R462Q) were downloaded from Human Genome Diversity Project (HGDP)-genotyped data (Li et al. 2008) (http://hgscl.org/hgdp/files.html). Age-standardized incidence rates (ASR) of prostate cancer in Europe, South Asia, Middle East, East Asia, North Africa, Oceania, Sub-Saharan Africa, and America were downloaded from GLOBOCAN 2008 (http://globocan.iarc.fr/). The Pearson correlation coefficient between allele frequencies and ASR was calculated to evaluate the association.

**Positively Selected Sites Detected by Likelihood Ratio Test**

In general, likelihood ratio tests implemented in the PAML package construct nested models with a general model and a corresponding model that allow for potential positively selected sites (Yang 1997). “Site-specific” models allowing $\omega = dN/dS$, the ratio of the nonsynonymous substitution rate ($dN$) to the synonymous substitution rate ($dS$), to vary among sites were used to detect candidate positively selected sites in the RNASEL gene (Nielsen and Yang 1998; Yang 2000). Here, we employed the nested models M1a versus M2a, models that are more conservative than others, where M2a is the model allowing positive selection. Significant difference between the two models was accepted if this difference was greater than twice the average log-likelihood difference between the models, a distribution that follows a $\chi^2$ distribution. When the likelihood ratio tests indicated significance, the Bayes empirical Bayes method was used to calculate posterior probabilities for the potential positively selected sites (Yang et al. 2005). Furthermore, the $dN/dS$ values on the different lineages were evaluated using a free-ratio model compared with a one-ratio model where all lineages have a single $dN/dS$ value with degrees of freedom (df) equal to the difference in the number of parameters of the two models. In addition, branch-site model was also used to detected signature of positive selection in the primate lineage (Zhang et al. 2005).

**Population Genetics Analysis**

Haplotypes were inferred with the PHASE program by Bayesian statistical methods based on the polymorphisms detected in the human populations (Stephens et al. 2001; Stephens and Donnelly 2003). Median-joining network for the inference of haplotype genealogy was constructed by Network 4.5.1.0 (Bandelt et al. 1999). Ancestral alleles for each SNP were deduced from the chimpanzee, and orangutan sequences obtained from UCSC (http://genome.ucsc.edu/).
Table 1. Likelihood Ratio Tests for Positive Selection on the RNASEL Gene.

<table>
<thead>
<tr>
<th>Lineages</th>
<th>Model</th>
<th>Null</th>
<th>Positive</th>
<th>Parameters</th>
<th>df</th>
<th>$\chi^2$</th>
<th>$P$ Value*</th>
<th>Positively Selected Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mammals</td>
<td>M1a versus M2a</td>
<td>-12292.11</td>
<td>-12249.05</td>
<td>2</td>
<td>86.12</td>
<td>$1.396 \times 10^{-18}$</td>
<td>40L*, 102A**, 137K*, 314H*, 3225*, 6105**;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M0 versus</td>
<td>-12515.45</td>
<td>-12515.45</td>
<td>30</td>
<td>72.01</td>
<td>$1.819 \times 10^{-04}$</td>
<td>639C*, 646K*, 651R**, 654F*, 735G*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M0 versus</td>
<td>-5298.90</td>
<td>-5279.41</td>
<td>19</td>
<td>38.99</td>
<td>$3.101 \times 10^{-02}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Branch-site model</td>
<td>-12270.74</td>
<td>-12270.74</td>
<td>1</td>
<td>42.74</td>
<td>$4.368 \times 10^{-10}$</td>
<td>137K*, 319V**, 379I*, 651R**, 735G*</td>
<td></td>
</tr>
</tbody>
</table>

*P values were adjusted by multiple testing.
*Denotes posterior probabilities 95% < $P$ < 99%.
**Denotes posterior probabilities $P$ < 99%.

Nucleotide diversity, $\pi$, and the proportion of segregating sites, $\theta_w$, were calculated for each human population. Traditional statistics including Tajima’s $D$ (Tajima 1989), Fu and Li’s $D$, $D^*$, $F^*$ (Fu and Li 1993; Fu 1997), and Fay and Wu’s $H$ (Fay and Wu 2000), used to detect the deviation from neutrality, were calculated by the program DnaSP 4.0 (Rozas et al. 2003). In addition, coalescent simulations were also constructed incorporating the best-fit human demographic parameters of Africans, Europeans, and East Asians, as described in Schaffner et al. (2005), to calculate the significance of the deviation from neutrality.

Results

Positive Selection on RNASEL among Mammals

Several sites in the sequenced RNASEL coding sequences of primates were found to be heterozygous. Ancestral alleles were deduced from the phylogenetic tree and were used for the analysis. A neighbor joining (NJ) tree of the 17 mammalian RNASEL coding sequences was constructed using Kimura 2–parameter distances with MEGA4.0 (Tamura et al. 2007), and this topology was used for likelihood ratio testing for positive selection (supplementary fig. 2, Supplementary Material online). In addition, an ML tree was constructed by PHYML using the HKY + Gamma substitution model with four substitution rate categories (Guindon and Gascuel 2003). The topology for the primate phylogeny found in the ML and NJ trees was identical. A total of 11 positively selected sites were identified in the RNASEL coding sequences by comparing models M2a against M1a in likelihood ratio tests (see Materials and Methods) (table 1), with five of these sites being located in the endoribonuclease domain, a number greater than that for any other protein region ($\chi^2 = 60.1, P = 0.014$ with df = 1), three located in the ankyrin repeats domain, and zero in the protein kinase domain (fig. 1). When the analysis was limited to only the 11 primate sequences, only 8 positively selected sites were identified. Site 541, which is associated with risk of developing prostate cancer, was not identified as being under positive selection by our likelihood ratio tests, a result that is inconsistent with a previous study (Summers and Crespi 2008), however that study had used only eight mammalian RNASEL sequences for their analysis and defined site 541 as positively selected site despite only having posterior probability $= 82\%$, whereas a conservative and rigorous analysis would have only defined it as being positively selected if the posterior probability was at least 95%.

We also calculated pairwise sequence distances, including the nonsynonymous ($dN$) and synonymous ($dS$) substitution rates by the Nei–Gojobori (Jukes–Cantor) method as implemented in MEGA4.0 (Tamura et al. 2007), for each of the three domains in the sequences among primates (fig. 2). The endoribonuclease domain had the highest evolutionary rate, with the protein kinase domain evolving at a much lower rate (fig. 2), indicative of the correlation between the functional role and evolutionary selective pressures with different functional regions evolving under divergent selective constraints. In RNASEL, the ankyrin repeats domain is responsible for receiving signals from the environment in combination with 2-5A, the protein kinase region is involved in enzyme dimerization by binding to ATP, and the C-terminal endoribonuclease

![Fig. 1. Mapping positively selected sites to the RNASEL protein. The domain structure of RNASEL is shown with the domains mapped by querying the Conserved Protein Database (Marchler-Bauer et al. 2005). Stars represent the 11 positively selected sites identified by M2a versus M1a in table 1.](https://academic.oup.com/mbe/article-abstract/29/10/3161/1030268?accessed=10%2FJanuary%202019)
domain catalyzes the degradation of pathogen RNA (Silverman 2003). The N- and C-terminal regions of RNASEL, thus, contact the virus directly, and the divergence in evolutionary rates for these regions supports a model of correlated evolution with an arms race between a host and its pathogen as proposed by the Van Valen’s (1973) Red Queen hypothesis.

A free-ratio model of substitution rates that assigns a different rate to each branch of the tree had a significantly better fit than a model with a single rate for all mammalian branches (table 1; \( \chi^2 = 72.01, P = 1.819 \times 10^{-04} \) with df = 30). This indicates that the \( dN/dS \) values for RNASEL vary among lineages. When we focused only on the primates, a similar conclusion was reached with different \( dN/dS \) values occurring among the different lineages (table 1; \( \chi^2 = 38.99, P = 3.101 \times 10^{-02} \) with df = 19). Branch-site model, with the primate lineage defined as foreground lineage, also found signatures of positive selection (table 1; \( \chi^2 = 42.74, P = 4.368 \times 10^{-10} \) with df = 1).

Polymorphisms in Humans and Incidence Rate of Prostate Cancer
As RNASEL is associated with prostate cancer (see Introduction); we also had an interest in the evolutionary pattern of RNASEL within humans. We sequenced RNASEL alleles from 144 individuals representing four separate populations including African, European, East Asian, and South Asian. A total of 26 mutations, including two deletions, were identified in the 4181 bp sequenced region. Both of the deletions were singletons and occurred in the coding regions: one of which resulted in deletion of amino acid 320L, whereas the other is a 21 bp (GTCTGCAGAAACGAGC- GAGA) deletion that results in a seven amino acid deletion and one amino acid change, and the 21 bp deletion was not used in the following analysis. Among the remaining 24 mutations, 12 are within the coding sequence with 7 being nonsynonymous and 5 synonymous substitutions. In a previous genome wide study by Bustamante et al. (2005), significant evidence for positive selection on the RNASEL gene was found using the McDonald–Kreitman test (McDonald and Kreitman 1991) by comparing human intraspecific polymorphisms with human–chimpanzee divergence. In our comparison, however, of nonsynonymous mutations to synonymous mutations between the human–chimpanzee divergence and human polymorphisms in 144 individuals, no significant evidence for positive selection was found (\( P = 0.71 \) by Fisher exact test).

To better understand the possible contributions of the nonsynonymous mutations, we used two methods, SIFT (Ng and Henikoff 2003) and PolyPhen (Ramensky et al. 2002) (table 2), to predict the functional effects of these amino acid changing substitutions. The amino acid changes (G59S and S689F) are predicted to be damaging by PolyPhen and SIFT and thus may have functional significance. G59S and S689F have very low derived allele frequencies, which is consistent with then being slightly deleterious mutations. A mutation that results in a termination codon at residue 265, and likely prevents the translation of a functional protein, was found in a heterozygous state in two individuals. Among our nonsynonymous substitutions, we identified two sites, R462Q and E541D, that previously been associated with prostate cancer (Carpten et al. 2002; Casey et al. 2002; Rokman et al. 2002; Li and Tai 2006; Robbins et al. 2008) and were predicted to be benign and should be tolerated, consistent with their high frequencies in modern humans.

For further analysis, we obtained the worldwide distributions of the SNP R462Q (fig. 3) and E541D (fig. 4), and the

![Fig. 2. Nonsynonymous substitution rate (\( dN \)) and synonymous substitution rate (\( dS \)) of the three domains of RNASEL among primate sequences. The pairwise sequence distances, including the nonsynonymous (\( dN \)) and synonymous (\( dS \)) substitution rates, were calculated by the Nei–Gojobori (Jukes–Cantor) method as implemented in MEGA4.0 (Tamura et al. 2007).](https://academic.oup.com/molbev/article-abstract/29/10/3161/1030268/1030268)

![Fig. 3. Worldwide distribution of allele frequency of SNP rs486907 (R462Q).](https://academic.oup.com/molbev/article-abstract/29/10/3161/1030268/1030268)

<table>
<thead>
<tr>
<th>Ancestral Allele</th>
<th>Derived Allele</th>
<th>PolyPhen</th>
<th>SIFT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position</td>
<td>Ancestral</td>
<td>Derived</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Allele</td>
<td>Allele</td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>G</td>
<td>S</td>
<td>Possibly damaging</td>
</tr>
<tr>
<td>97</td>
<td>I</td>
<td>L</td>
<td>Benign</td>
</tr>
<tr>
<td>265</td>
<td>E</td>
<td>X</td>
<td>N/A</td>
</tr>
<tr>
<td>462</td>
<td>R</td>
<td>Q</td>
<td>Benign</td>
</tr>
<tr>
<td>541</td>
<td>E</td>
<td>D</td>
<td>Benign</td>
</tr>
<tr>
<td>565</td>
<td>R</td>
<td>H</td>
<td>Benign</td>
</tr>
<tr>
<td>689</td>
<td>S</td>
<td>F</td>
<td>Possibly damaging</td>
</tr>
</tbody>
</table>

Table 2. PolyPhen and SIFT Prediction of Amino Acid Polymorphisms.
age-standardized incidence rates (ASR) of prostate cancer for Europe, South Asia, Middle East, East Asia, North Africa, Oceania, Sub-Saharan Africa, and America. The incidence of prostate cancer in Asians is much lower than in Western countries, including Oceanian, North American, and European countries. Several nongenetic factors might contribute to the pattern seen in Asia, such as nutrition, where soy foods are popular among Asians, especially nonfermented soy foods, which have consistently been reported to be associated with a 25–30% reduction in the risk of prostate cancer (Kurahashi et al. 2007; Hwang et al. 2009; Yan and Spitznagel 2009) (reviewed in Kimura 2011). Herein, after excluding the data from East Asians, the frequency of the 541D allele was significantly associated in a negative direction with the age-standardized incidence rate across the World ($P = 0.0259$) (fig. 5B), as was the frequency of the 462R–541D haplotype ($P = 0.0243$) (fig. 5C). These observations are consistent with the 541D allele being associated with a decrease in the risk for prostate cancer (Robbins et al. 2008) and indicated that genetic factor significantly contribute to the difference in incidence rates of prostate cancer among different populations of the World.

**Population Variation of RNASEL in Humans**

A total of 29 haplotypes were inferred based on the 25 polymorphisms using the program PHASE (Stephens et al. 2001; Stephens and Donnelly 2003). Population statistics, including nucleotide diversity, $θ_w$, Tajima's $D$, Fu and Li's $D$, $D^*$, $F^*$, and Fay and Wu's $H$, are summarized in supplementary table 1 (Supplementary Material online). A median-joining network of the haplotypes was also constructed (fig. 6).

Interestingly, haplotypes carrying the derived allele D (Aspartic acid) of SNP rs627928 (E541D) demonstrated a lower level of nucleotide diversity ($θ_w = 2.32 \times 10^{-4}$) and lower value of Tajima's $D_T$ value ($-1.85$, $P < 0.05$, and also demonstrates statistical significance after coalescent simulations incorporating the best-fit human demographic parameters, see supplementary table 2, Supplementary Material online) than haplotypes carrying the ancestral allele E ($θ_w = 3.94 \times 10^{-4}$, and Tajima's $D_T$...
For SNP E541D is not

The lower $D_T$ value for SNP E541D is not due to negative selection since it has a high derived allele frequency in human populations (49.31%). In addition, interestingly, an excess of haplotypes, with lower frequencies, were observed to have a star-like relationship with 541D in the haplotype genealogy (fig. 6), with most of these low frequencies haplotypes belonging to Africans. The derived allele (i.e., D) frequency of E541D reaches 72.06% in the African populations, the population that has the lowest value for Tajima’s $D$. This result is suggestive that positive selection probably operated on the RNASEL gene in Africans. In contrast, the Tajima’s $D_T$ for haplotypes carrying the ancestral allele R of the SNP rs486907 (R462Q) demonstrated a lower value than those carrying the derived allele but was not statistical significant (table 3; supplementary table 2, Supplementary Material online).

Haplotypes composed by the two amino acid polymorphic sites: R462Q and E541D were Q-E, R-E, and R-D and had 74, 71, and 143 chromosomes, respectively. Intriguingly, the Q-D haplotype was not observed in any of the individuals tested, thus all individuals with the 541D haplotypes are R-D (table 3). The Tajima’s $D$ value for the R-D haplotypes is $-1.85$ ($P < 0.05$), and thus, the high frequency for R-D (49.65%) was probably driven by positive natural selection, with the target of the selection likely being the derived 541D allele rather than the ancestral 462R allele.

**Discussion**

The N-, and C-terminal domains of RNASEL, domains that are responsible for receiving environmental signals and digesting viral RNA were found to evolve more rapidly than the kinase domain among mammals. The more rapid evolution of these domains is consistent with the Van Valen’s (1973) Red Queen hypothesis that there is an arms race between host and pathogen; therefore, our evolutionary analysis provides evidence of a correlation between function and evolution for the RNASEL gene.

Prostate cancer is a cancer of the prostate, a gland involved in the male reproductive system. Epidemiological studies have indicated that infection and inflammation likely contribute to the development of prostate cancer (De Marzo et al. 2007). Many different pathogenic organisms have been observed to infect and induce an inflammatory response in the prostate, including *Neisseria gonorrhoeae*, *Chlamydia trachomatis*, *Trichomonas vaginalis*, *Treponema pallidum*, and *Propionibacterium acnes*, as well as some viruses such as human papillomavirus (HPV), human herpes simplex virus type 2 (HSV2), cytomegalovirus (CMV), and human herpes virus type 8 (HHV8) (De Marzo et al. 2007).

A previous study has reported that the haplotype 462R–541D shows a strong protective effect against prostate cancer (Sen et al. 2009). One further possibility is that positive selection could explain the higher frequency of the derived allele 541D in African populations.
Figure 7. Frequencies of haplotypes of the two SNPs 364R/Q and 541D/E in the worldwide panel. The haplotypes of SNPs in RNASEL from the Human Genome Diversity Cell Line Panel were inferred with the PHASE program (Stephens et al. 2001; Stephens and Donnelly 2003).

Cancer in African Americans (Robbins et al. 2008). We propose that the protective effect of RNASEL with the 462R–541D substitutions is due to a higher effective activity of the enzyme defending against infections by pathogens, thus reducing the initiation of prostate cancer. Genotypes at SNP rs627928 (E541D) and SNP rs486907 (R462Q) were obtained from a worldwide panel in HGDP (Li et al. 2008), and the haplotype R-D was confirmed to have high frequency in all African populations (~50–76%) (fig. 7). Our study suggests that the RNASEL haplotype 462R–541D (e.g., 541D) probably evolved due to positive selection in Africans, due to the advantage of this haplotypes at defending against infection by viruses. Previous studies found that many cancer-related genes are under positive selection, which was explained by them having adaptive benefit for one aspect of biology (e.g., defending against pathogen infections) but having a pleiotropic effect in increasing the risk of cancer (Crespi and Summers 2006). In contrast, here, we found that positive selection on variants of RNASEL defended against infection by viruses and decreased the prostate cancer risk. Our data present a link among positive selection of host genes, infection by virus, host defense, and cancer. It has been established that infection by some pathogens can cause cancers (De Marzo et al. 2007) and that some genes for pathogen defense are associated with the cancers (fig. 8). Here, we show that an arms race occurred between host genes and a virus during evolution that drove positive Darwinian selection to operate on some genes in the host that play a role in the defense against viruses that resulted in a reduced risk of cancer (Van Valen 1973).

Figure 8. Links among positive selection, infectious viruses, host, and cancer in the host: An arms race occurs between the host and virus resulting in positive Darwinian selection operating on some variants at the genes that play roles in the defense against viruses; infection by some pathogens could cause cancer in the host; and some variants at the genes for defense are associated (e.g., decrease the risk) with cancers.

Supplementary Material
Supplementary materials 1 and 2, figures 1 and 2, and tables 1 and 2 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments
Thanks to Dr. Joseph K. Pickrell for providing his program to draw the worldwide distribution of allele frequencies. This work was supported by grants from the National Natural Science Foundation of China (31061160189).

References


