Journal of Molecular Evolution manuscript No.
(will be inserted by the editor)

Detecting site-specific physicochemical selective pressures

Applications to the class-I HLA of the human major histocompatibility complex and the SRK of the plant sporophytic self-incompatibility system

Raazesh Sainudiin[1], **Wendy Shuk Wan Wong**[2], **Krithika Yogeeswaran**[3], **June B. Nasrallah**[3], **Ziheng Yang**[4], **Rasmus Nielsen**[2]

[1] Department of Statistical Science, Cornell University, Ithaca, NY 14853
[2] Department of Biological Statistics and Computational biology, Cornell University, Ithaca, New York 14853
[3] Department of Plant Biology, Cornell University, Ithaca, New York 14853
[4] Department of Biology, University College London, London, UK

The date of receipt and acceptance will be inserted by the editor

Abstract Models of codon substitution are developed that incorporate physicochemical properties of amino acids. When amino acid sites are inferred to be under positive selection, these models suggest the nature and extent of the physicochemical properties under selection. This is accomplished by first partitioning the codons on the basis of some property of the encoded amino acids. This partition is used to parametrize the rates of property-conserving and property-altering base substitutions at the codon level by means of finite mixtures of Markov models that also account for codon and transition:transversion biases. Here, we apply this method to two positively-selected receptors involved in ligand-recognition: the class-I alleles of the human Major Histocompatibility Complex (MHC) of known structure and the S-locus Receptor Kinase (SRK) of the sporophytic self-incompatibility system (SSI) in cruciferous plants (*Brassicaceae*), whose structure is unknown. Through likelihood ratio tests we demonstrate that at some sites, the positively selected MHC and SRK proteins are under physicochemical selective pressures to alter polarity, volume, polarity &/or volume and charge to various extents. An empirical Bayes approach is used

Correspondence to: Raazesh Sainudiin, Department of Statistical Science
301 Malott Hall, Cornell University Ithaca, NY 14853, USA
Phone: (607) 255-8066 **Fax**: (607) 255-9801 **email**: rs228@cornell.edu

to identify sites that may be important for ligand recognition in these proteins.

Key words Codon-based Markov models – likelihood ratio tests – MHC – physicochemical selective pressures – SRK

1 Introduction

The extent to which an amino-acid (AA) residue may freely change depends on its structural and functional role within a protein. The neutral theory of molecular evolution posits that most of the observed polymorphism at the molecular level is due to the random fixation of selectively neutral mutations (Kimura, 1983). At some codon sites, nonsynonymous (AA-altering) mutations are not tolerated due to strong purifying selection, while at other sites only a particular subset of nonsynonymous mutations that preserve the function of the protein by preserving its overall structure are tolerated (Pakula and Sauer, 1989). Therefore, selectively neutral point mutations at a codon site that ensure a protein's functionality may be of two types: all synonymous (AA-conserving, structurally-silent, and hence functionally-silent) mutations and certain tolerable nonsynonymous mutations (AA-altering, not structurally-silent, but functionally-silent). In other words, a neutrally evolving AA site is likely to remain unchanged or preferentially change over time. Thus, the rate of neutral evolution is not always equal to the rate of synonymous substitutions alone. It has been well established that at such preferentially changing sites, AA substitutions among physicochemically similar amino acids are more frequent than those between dissimilar ones (Zuckerkandl and Pauling, 1965; Sneath, 1966; Epstein, 1967; Clarke, 1970; Grantham, 1974; Miyata et al., 1979).

Highly variable regions of rapidly-evolving proteins that are functionally important are usually targeted by natural selection. They tend to have an excess of nonsynonymous substitutions compared to what would be expected if both synonymous and nonsynonymous substitutions occur at the same rate. Under the assumption that neutral evolution is reflected by a nonsynonymous to synonymous substitution rate ratio of 1, codon-based statistical methods exist in the literature to test for the presence and quantify the strength of positive selection. Such methods can model rate heterogeneity among sites, transition:transversion and codon biases operating at the DNA level, and rate heterogeneity among codon substitutions. In order to model heterogeneity in rates between distinct pairs of nonsynonymous codons, Yang et al. (1998) assume a parametric form based on an amino acid metric of Miyata et al. (1979), while Schadt and Lange (2002) partition the amino acids into similarity classes and parametrize acceptance probabilities for transitions within or between classes. Such methods shed light on the nature and strength of positive selection by deciphering any physicochemically meaningful patterns in nonsynonymous substitutions. We build on the codon-based models of Nielsen and Yang (1998) to provide a likelihood framework to detect an elevation in the rate ratio of property-altering to property-conserving substitutions. The empirical Bayes method of Nielsen and Yang (1998) is further used to compute the posterior probability that a particular site is subject to an elevated rate ratio (> 1). Our method can be applied to any physicochemical property of interest.

We show that HLA-A and HLA-B, the class-I glycoproteins involved in peptide recognition in the human major histocompatibility complex (MHC), as well as, SRK, the female receptors involved in recognizing self-pollen in the sporophytic self-incompatibility system (SSI) of cruciferous plants, are positively selected at some sites as reflected by a high rate ratio of nonsynonymous to synonymous substitutions. We explore the physicochemical nature of this positive selection through the reflections of elevated rate ratios of property-altering to property-conserving substitutions at some sites. All the investigated properties, namely, polarity, volume, polarity &/or volume, and charge, show elevated rate ratios at some sites. Specific residues of these proteins that may be the targets of positive selection for changes in any of these physicochemical properties are identified and surmised to be important for ligand recognition.

2 Theory

The methods of Nielsen and Yang (1998) and Yang et al. (2000) are based on modeling the evolution of the coding sequence as a continuous time Markov chain with state space on the set of sense codons. The rate of a nonsynonymous nucleotide substitution on a codon at site h relative to that of a synonymous substitution at any site is parametrized as $\omega^{(h)} \in (0, \infty)$. The transition/transversion rate ratio κ and the stationary frequency π_v of codon v model the mutaional biases at the DNA level. In order to allow rate variation along the protein sequence, $\omega^{(h)}$ is considered unknown and drawn from some finite mixture of rate classes with possibly distinct rates. The likelihood function is calculated via the pruning algorithm of Felsenstein (1981) subsequent to the superimposition of the codon substitution process along the branches of the phylogenetic tree (Nielsen and Yang, 1998). To test if there is any evidence for positive selection, a likelihood ratio test may be performed between model M7 and model M8. We use the M7-M8 likelihood ratio test as it has been shown by Anisimova et al. (2003) to be more robust in the presence of recombination (see Discussion section). Model M7 uses a discretized Beta distribution taking values between 0 and 1 in order to draw values for ω, and model M8 draws ω from a mixture of a discretized Beta distribution (as in M7) and an extra rate class that is allowed to be > 1 (Yang et al., 2000). If ω is > 1 at a site, then that site is thought to be under positive selection. Rejecting M7 is interpreted as rejecting the hypothesis that no sites are undergoing positive selection. Furthermore, if M7 is rejected, then the positively selected sites can be detected. This is accomplished by computing the posterior probability that a particular site has a value of $\omega > 1$ under the parameter estimates of the M8 model through an empirical Bayes approach (Nielsen and Yang, 1998). Since the M7-M8 likelihood ratio test statistic has a non-trivial asymptotic distribution due to nonstandard conditions, we use the χ^2_2 distribution to conservatively compare M7 and M8 models (Anisimova et al., 2001).

In the above formulation, the state space of sense codons is partitioned into codons that code for the same AA. If the modeling aspects of transition:transversion and codon biases (κ and π_v's) are ignored for the moment, then the substitution rate between codons that code for the same AA will be 1 for all sites and that between codons that code for different AAs will be ω. In other words, all else being equal, the nonsynonymous substitution rate is scaled relative to the synonymous substitution rate. A simple generalization of the above framework that allows one to explore the nature of different selective pressures acting on a protein is proposed in this study. Instead of partitioning the codons on the basis of the encoded AA, they can be partitioned according to some physicochemical property. For instance, if polarity were chosen to be the property of interest, the sense codons would be partitioned into those that code for polar amino acids and those that do not. The polarity-conserving rate (*i.e.* the rate of substitution between codons that code for polar amino acids and that between codons that code for nonpolar AAs) is set at 1. The polarity-altering rate (*i.e.* the rate of substitution between a codon that encodes for a polar AA and another that encodes for a nonpolar AA, and vice-versa) is defined to be γ_p. One may model γ_p to vary among sites through mixture models as before. Let model M7p and M8p under the polarity-based partition of the codons be the analogs of models M7 and M8 of Yang et al. (2000). If model M7p under the polarity partition is rejected in favor of model M8p, then one may similarly conclude that there is evidence in favor of an elevated rate of polarity-altering substitutions compared to the rate of polarity-conserving substitutions at some sites. Subsequently, site-specific posterior probabilities of being subject to an elevated polarity-altering rate when compared to the polarity-conserving rate ($\gamma_p > 1$) can be computed through an analog of the empirical Bayes approach of Nielsen and Yang (1998).

In general, by specifying a partition of codons based on some set of properties of the encoded AA, one can gain insight into the nature of specific physicochemical selective pressures acting upon a protein. The rate matrix Q for such a codon substitution process is given by,

$$q_{uv} = \begin{cases} \pi_v \ , & \text{if } u \text{ and } v \text{ differ by a property-conserving transversion} \\ \kappa \ \pi_v \ , & \text{if } u \text{ and } v \text{ differ by a property-conserving transition} \\ \gamma^{(h)} \ \pi_v \ , & \text{if } u \text{ and } v \text{ differ by a property-altering transversion} \\ \gamma^{(h)} \ \kappa \ \pi_v \ , & \text{if } u \text{ and } v \text{ differ by a property-altering transition} \\ 0 \ , & \text{if } u \text{ and } v \text{ differ at more than one position.} \end{cases} \tag{1}$$

Observe that property-conserving substitutions include synonymous substitutions. The parameter κ is the transition/transversion rate ratio and π_v is the stationary frequency of codon v. Note that $\gamma^{(h)} = \omega^{(h)}$ if the state space is partitioned into codons that encode the same AA.

Suppose $\mathcal{A} = \{A_1, A_2, \cdots, A_i, \cdots, A_m\}$ is a partition of the set A of 20 amino acids, where $m \in \{1, \cdots, 20\}$ and $A_i = \{a_{i1}, a_{i2}, \cdots, a_{iz_i}\}$ denotes the set of z_i distinct amino acids. Each partition $\mathcal{A}$ of the amino acids induces a corresponding partition $\mathcal{C}$ of the 61 sense codons, such that $\mathcal{C} = \{C_1, \cdots, C_m\}$, where $C_i = \bigcup_{j=1}^{z_i} c_{ij}$ and c_{ij} is the set of codons that code for amino acid $a_{ij} \in A_i$. Thus C_i is the set of codons which code for the amino acids in the set A_i. This partition $\mathcal{C}$ induces a Markov chain model which parametrizes $\gamma_{\mathcal{A}}$ as the property-altering substitution rate between codons $u \in C_i$ and $v \in C_j$, where $i \neq j$, relative to the property-conserving rate between codons $u, v \in C_i$ for all i. There are more than 6×10^{18} models that correspond to the number of ways one can partition the set of amino acids. Out of such a large class of models, some models with physicochemically meaningful partitions are of interest.

In receptor-ligand interactions, the two proteins come in intimate contact with each other to allow recognition. At this interface, the side-chain residues from one protein interact with their counterparts on the other to form hydrogen bonds, salt bridges, etc. in physicochemically feasible ways. Tight contact is ensured when complementary residues accomodate each other by their size and shape and when repulsion due to similar charge and/or dissimilar polarity are avoided (Creighton, 1996). For these reasons, we study models induced by each of the four partitions described in Table 1 in addition to the standard model based on the ω rate ratio. Models induced by these partitions are used to detect selective forces driving changes in these properties. One may similarly look at models with other partitions, such as, hydrophobicity, aromaticity, and electrostatic potential. Furthermore, appropriate partitions may be designed in light of empirical evidence for the particular system under investigation to test the extent of the hypothesized physicochemical pressures targeted by positive selection.

[Table 1 about here.]

3 Materials and Methods

Gene name, source organism, and GenBank accession number for each of the analyzed MHC class I HLA genes and S-locus receptor protein Kinases are given in Table 2. Amino acid sequences from each data set were aligned with TCOFFEE (http://www. ch.embnet.org/software/TCoffee.html). The multiple alignment of AA sequences was used to obtain the corresponding codon alignment (http://bioweb.pasteur.fr/seqanal/interfaces/ protal2dna-simple.html). The maximum likelihood tree topology from the nucleotide sequences for each data set was obtained using the DNAML program (version 3.5c) in PHYLIP (http:// evolution.gs.washington.edu/phylip.html).

We obtained a 95% confidence set of trees for SRK under a Bayesian framework using PHYBAYES (http://statgen.ncsu.edu/ stephane/softs.htm).

[Table 2 about here.]

First, a likelihood ratio test is performed by comparing models M7 and M8 under the standard AA-based partition to detect an elevation in ω, the nonsynonymous to synonymous rate ratio, for each data set. If the null model M7 is rejected in favor of model M8, then there is strong evidence against the ω rate ratio being confined to the interval $[0, 1]$ for all sites. This is interpreted as rejecting the absence of positive selection at all sites in favor of its presence at some of the sites. Second, we investigate the extent to which various physicochemical properties are targeted by positive selection in homologous proteins evolving with a high ω rate ratio. This is accomplished by performing similar likelihood ratio tests. Four such tests, namely, M7p versus M8p, M7v versus M8v, M7pv versus M8pv, and M7c versus M8c, are performed. These four pairs of models are induced by the partitions based on polarity, volume, polarity &/or volume, and charge, respectively, as described in Table 1. If the null model is rejected in each of these four tests, then there is evidence for an elevation in the property-altering substitution rate relative to the property-conserving substitution rate for polarity (γ_p), volume (γ_v), polarity &/or volume (γ_{pv}), and charge $9\gamma_c$), at least at a few sites. The source code of the *codeml* program in PAML (http://abacus.gene.ucl.ac.uk/software/paml.html) was modified to accomplish the task. When the null model is rejected we use the empirical Bayes approach described earlier to obtain the posterior probability that a given site has a ratio of the property-altering substitution rate to the property-conserving substitution rate > 1. If this posterior probability is > 0.95, then we say that the site is under pressure to change property $\mathcal{A}$ or that $\gamma_{\mathcal{A}} > 1$ at the site.

The likelihood analyses for each data set were done only on one tree topology, namely the maximum likelihood topology. In order to ensure that uncertainty in tree topology does not affect the conclusions drawn from hypothesis tests, the likelihood ratio tests were repeated on each of 17 distinct but similar topologies that contained 95% of the posterior probability mass on the space of possible topologies for 21 SRKs using PHYBAYES. Rejection of model M7 in favor of M8 occurred under each topology (results not shown). Simulations were used to evaluate the performance of the posterior predictions that a given site is under a certain physicochemical pressure. Each of the two codon models, induced by partitions based on polarity and charge, respectively, was used to simulate 1000 data sets of homologous codon sequences under the maximum likelihood parameter estimates of M8p and M8c models, respectively, for the SRK data. The posterior predictions were conducted on the simulated data and the false positive rates were found to be extremely low ($< 1\%$).

4 Applications

4.1 MHC class-I:

Proteins encoded by genes at any one of the three highly polymorphic MHC class I loci, HLA-A, HLA-B, and HLA-C, are expressed on the cell surface to present peptides to T-cell receptors (TCRs). This presentation leads to a cascade of cellular immune responses culminating in the destruction of the infected cell. The polymorphism is believed to be maintained by selection for resistance to pathogens evolving to escape immune detection by avoiding HLA presentation to T-cells (Doherty and Zinkernagel, 1975). The structure of HLA is well understood (Madden, 1995). The peptide binding region (PBR) is a highly polymorphic cleft formed between two α-helices lying on a β-pleated sheet. The side chain residues directed towards the interior of this cleft form ridges that demarcate six peptide-binding pockets (A-F). All the pockets, especially B-E, contain variable residues that determine the specificity of the interaction by determining the size and shape of the pockets and thereby the size and shape of side-chain residues of the peptide that can be accomodated. It has been reported that even a small number of amino acid changes in the cleft will result in drastic changes in the types of peptides the HLA protein can bind to (Barber et al., 1997). Amino acid replacements in this cleft appear to be driven by balancing selection from several lines of empirical evidence (Hedrick and Thomson, 1983; Hughes and Nei, 1988; Hedrick et al., 1991; Markow et al., 1993; Black and Hedrick, 1997). The six class I MHC alleles from HLA-A and HLA-B loci chosen for this study were already shown by Swanson et al. (2001) to be under positive selection as reflected by an elevated nonsynonymous to synonymous substitution rate ratio at some sites. We further analyze this data set with models induced by other physicochemical partitions in order to gain insight into the nature and extent of various physicochemical properties targeted by positive selection at each site.

[Table 3 about here.]

The results of the likelihood ratio tests to detect site-specific elevations in ω, γ_p, γ_v, γ_{pv}, and γ_c of the MHC are shown in Table 3. In each of the five tests, the null model is strongly rejected. The amino acid sites of the MHC protein that have a high posterior probability (> 0.95) of being under at least one of the elevated substitution rate ratios are identified. We find that under this stringent posterior probability cut-off, $\omega > 1$ at sites 114 and 156, $\gamma_p > 1$ at site 116, $\gamma_v > 1$ at sites 63, 67, and 97, $\gamma_{pv} > 1$ at sites 45, 63, 67, and 97, and $\gamma_c > 1$ at sites 45, 114 and 156. These sites are mapped onto the crystal structure of an MHC class I HLA-A2 protein (PDB file 1QSE) using RASMOL v2.7.2.1 (http://www.bernstein-plus-sons.com/software/rasmol/) as shown in Figure 1. All 7 selected sites are found to be present in one or more of the 6 peptide-interacting pockets (A-F) of the antigen binding cleft. Residues 45 and 67 are found in Pocket

B, 63 in A and B, 97 in C and E, 114 in C, D and E, 116 in C and F and 156 in D and E.

[Fig. 1 about here.]

Approximately 1.5% of all sites in the MHC proteins are under polarity-altering pressure ($\gamma_p > 1$), whereas 6%, 7%, and 14% are under charge-altering ($\gamma_c > 1$), AA-altering ($\omega > 1$), and volume-altering ($\gamma_v > 1$) pressures, respectively. Thus, different proportions of sites are under different kinds of physicochemical selection pressures. The small proportion of sites with $\gamma_p > 1$ in the MHC is consistent with the observation that most non-synonymous codon substitutions involving a single mutation are polarity-conserving (Epstein, 1967). The larger proportion of sites with $\gamma_v > 1$ may partly reflect that substitutions altering the volume of an amino acid in pockets B-E of the antigen-binding cleft more often improve its ability to physically accomodate novel peptides, compared to substitutions that alter other properties, and are thereby selected and retained in the population.

Out of the 43 polymorphic sites in our alignment of 6 MHC proteins, only 7 sites are seen to be targeted by positive selection as reflected by an elevation in at least one of the five substitution rate ratios under study. All these sites are within an Ångstrom from active Van der Waal interactions with a peptide residu based on an analysis of several structurally resolved peptide-MHC complexes (Reche and Reinherz, 2003). Though no structural studies on the 6 HLA variants selected for this study exist, x-ray crystallographic studies of other HLA-A and HLA-B proteins (Guo et al., 1993; Macdonald et al., 2003), site-specific mutagenesis studies (Domenech et al., 1991) and studies on varying specificity resulting from single site variations in HLA-B subtypes (Hulsmeyer et al., 2002; Macdonald et al., 2003) report all 7 of these selected sites (or regions containing them) to be critical in peptide recognition and peptide repertoire determination.

The only two sites, 114 and 156, with a high posterior probability of $\omega > 1$ are located in Pocket E. Site 156 was determined to be buried at the base of this pocket, but still critical in influencing the extent of cleft opening through its interactions with adjacent residues in HLA-B44 variants (Macdonald et al., 2003). D156 (negatively-charged) forms a salt bridge with R97 (positively-charged) which allows a H-bond with D114, a strong interaction that indirectly results in the tightening of the whole cleft. When AA156 is neutral, a shallow pocket is presented for residue-3 of the ligand to interact with. We found that sites 114 and 156 were under pressure to alter charge ($\gamma_c > 1$) but not polarity, while site 97 was found to have the highest posterior probability of being targeted for changes in volume ($\gamma_v > 1$). Each of the 5 other critical sites, namely 45, 63, 67, 97, and 116, has a posterior probability > 0.80 that $\omega > 1$. Clearly, these sites have escaped detection under the standard AA-based partition due to the stringent cut-off of 0.95 which was chosen to increase accuracy in the possible presence of any unaccounted recombination (see Discussion section).

[Fig. 2 about here.]

Sites 45, 63 and 67 all lie within pocket B which is critical for specificity. The side chain of residue 2 of the peptide (P2) interacts with these residues in this pocket. In B*3501, F67 forms a pocket sterically favorable for a proline at P2 (Smith et al., 1996). In two HLA-A alleles, V67 may either block or permit access to the bottom of the pocket depending on the orientation of its side chain. In the former case this leads to selection of small side-chained P2 residues (A*6801), and in the latter case, to the selection of larger, non-polar side chains at site P2 to interact with the non-polar M45 in A*0201 (Guo et al., 1993). Site 67 is found to be selected for changes in volume ($\gamma_p > 1$) and polarity and/or volume ($\gamma_{pv} > 1$) concordant with these findings. Site 45 is located at the bottom of the B pocket. This site is a conserved M in all HLA-A alleles but variable in HLA-B alleles (Reche and Reinherz, 2003). In the latter case, variants observed include E (negatively-charged) and K (positively-charged) which select for R (positively-charged) and E (negatively-charged), respectively at site P2 in the peptide resulting in the formation of salt bridges between them (Macdonald et al., 2003; Guo et al., 1993), aside from neutral residues like T and M (Reche and Reinherz, 2003). We observe this site to be targeted for changes in charge ($\gamma_c > 1$) in our analysis. Since this site also has an elevated substitution rate ratio under the polarity and/or volume partition ($\gamma_{pv} > 1$), both these properties may be critical in residue selection at P2, especially for HLA-A and others B alleles with neutral AA residues at this position.

Site 116 is the only site with an elevated rate ratio exclusively under the polarity partition ($\gamma_p > 1$). This site is located at the bottom of the F Pocket and is important in determining ‘anchoring residues’ in the C-terminal end of the peptide ligand. Polarity appears to be critical in allele HLA-B*4402 where D116 (polar) forms a direct H-bond to Y74 resulting in the tightening of the entire binding cleft (Macdonald et al., 2003). Selection for change in polarity at site 116 may therefore be pivotal in tightening or relaxing the binding cleft and consequently narrowing or expanding the range of peptide specificity. However, studies in mammals demonstrate that bulky aromatic residues (Y or F) that select L, I or V in the peptide, or small residues (D or S) which accomodate the long side chain-bearing peptide residues (R or K), are commonly seen at this position (Young et al., 1995; Falk et al., 1991). Therefore, in this particular case polarity does not seem to be the only property that might be subject to selective pressures, but a potentially crucial one, at least for this data set. A closer look at our data shows a posterior probability of 0.88 for this site under the volume-altering partition, below the threshold of 0.95. It is helpful to look at the posterior probabilities under all five partitions for the sites selected by at least one partition (Figure 2).

4.2 SRK:

The self/nonself discriminating sporophytic self-incompatibility (SSI) system of the crucifer family (*Brassicaceae*) prevents self-fertilization and promotes out-crossing (see Nasrallah (2002) for a review). Specificity in SSI is determined by two highly polymorphic genes that are tightly linked within the *S*-locus complex. One gene encodes the *S*-locus receptor kinase protein (SRK), a single-pass transmembrane serine/threonine kinase displayed on the stigmatic surface of the female pistil (Stein et al., 1996). The second gene encodes the *S*-locus cysteine-rich protein (SCR), a 50-59 amino-acid long, hydrophilic, and positively charged protein located in the outer coat of pollen grains (Schopfer et al., 1999). SCR has been shown to be the ligand for SRK, and specificity in the SSI response results from allele-specific interactions between the SRK and SCR (Kachroo et al., 2002). During pollination, pollen grains released from the male anthers come in contact with the stigma, allowing the SCRs to interact with the SRKs. Upon self-pollination, SCR binds the SRK ectodomain, thereby activating the receptor and triggering a signalling cascade that leads to inhibition of pollen tube development at the stigma surface. The implication of this mechanism of self-recognition is that the SRK and SCR genes must co-evolve to maintain the specific interaction of their products.

Crucifer species typically exhibit a large number of *S* haplotypes, each with unique *SRK* and *SCR* alleles. Several of these *S* haplotypes predate speciation events (Dwyer et al., 1991; Uyenoyama, 1995) and are thought to be under balancing selection. Nishio and Kusaba (2000) have established three contiguous hypervariable regions (HVR1, HVR2, HVR3) and a C-terminal variable region (CVR) in the SRKs using approximate methods. By computing the average numbers of synonymous (π_S) and nonsynonymous (π_N) nucleotide differences per site between two randomly chosen sequences, Sato et al. (2002) show that $\pi_N{:}\pi_S > 1$ in the hypervariable regions indicating positive selection. In the absence of empirical information, a sliding window approach is used to decide the candidate hypervariable regions. Such candidate regions are typically contiguous stretches of the primary sequence which may not necessarily contain all sites that function together in ligand recognition in the tertiary structure of the receptor. Approximate methods are also available to compute the average pair-wise ratio of the proportion of radical nonsynonymous substitutions over the proportion of conservative nonsynonymous substitutions (p_{NR}/p_{NC}) (Nei and Gojobori, 1986; Hughes et al., 1990; Zhang, 2000). However, when measures of positive selection do not simultaneously account for the genetic code, codon and transition:transversion biases, and AA composition, they have been shown to be sensitive to mutational and compositional factors (Dagan et al., 2002).

[Fig. 3 about here.]

[Table 4 about here.]

Likelihood ratio tests identical to those performed with the MHC are carried out on the SRK proteins and summarized in Table 4 and the maximum likelihood tree under the M8 model for the 21 SRK sequences appears in Figure 3. Due to a scaling of the rate matrix Q, the branch lengths of this tree measure the expected number of nucleotide substitutions per site and the estimates of the parameters in Q capture the pattern of substitution. The total length of this estimated tree is 5.05. All five null hypotheses are strongly rejected for the SRK data set as well. Thus, there is strong evidence against the absence of positive selection at all sites, in terms of the substitution rate ratios, ω, γ_p, γ_v, γ_{pv}, and γ_c induced by their respective partitions being confined to the interval $[0, 1]$ for all sites. Figure 4 shows all 40 amino acid sites with a high posterior probability (> 0.95) that at least one of these substitution rate ratios is > 1. The numbering corresponds to the protein sequence of *B. oleracea* SRK60. Intuitively, the different partitions of the codon space and the corresponding Markov models they induce may be thought of as different sieves, each efficient at filtering out sites with certain patterns of substitutions relative to certain other patterns of substitution. Thus, by using several sieves one can filter out several kinds of sites. Figure 4 would thus show the results of several sieves applied to filter out SRK sites under different physicochemical pressures.

[Fig. 4 about here.]

All these selected sites lie in the S domain of SRK which functions in recognition of self-SCR. Many of the selected sites are under all five selective pressures. It is likely that these sites which allow flexibility in the nature of tolerated nonsynonymous mutations, contribute exclusively to ligand recognition rather than to the structural integrity of the protein. However, some sites are only under specific combinations of pressures. For instance, sites 206, 217, and 105 all have high posterior probabilities (> 0.95) of being under AA-altering pressure ($\omega > 1$). However, only the first two sites have high posterior probabilities of being under polarity-altering pressure ($\gamma_p > 1$), whereas site 105 is under volume-altering pressure ($\gamma_v > 1$). In contrast to sites under all five pressures, these sites are perhaps more structurally constrained and thus allow for property-specific nonsynonymous substitutions that may facilitate ligand recognition. Sites 208 and 215 of SRK have posterior probabilities > 0.95 of being under AA-altering ($\omega > 1$) and charge-altering ($\gamma_c > 1$) pressures. This information about these positively selected sites being exclusively targeted for charge alteration is useful since the coevolving SCR is a highly-charged ligand. Several sites that did not have a posterior probability > 0.95 of $\omega > 1$ have high posterior probabilities of being under an elevated property-altering to property-conserving rate ratio (sites outside the solid ellipse in Figure 4). Each of these 40 selected sites is subject to different physicochemical pressures to varying extents (Figure 5).

[Fig. 5 about here.]

Sites 205-220, 270-306, 328-343, and 413-423 of *B. oleracea* SRK60 correspond respectively to the hypervariable regions HVR1, HVR2, HVR3, and CVR established by Nishio and Kusaba (2000), which together comprise 80 sites in these four hypervariable regions. Our analysis presents 40 sites with high posterior probabilities of being under at least one of the five physicochemical selective pressures studied. Out of these 40 sites, only 27 belong to the hypervariable regions of Nishio and Kusaba (2000). The remaining 13 sites, shown in bold italics in Figure 4, may be vital for ligand recognition in the tertiary structure. We found 14 out of the 40 sites to have lower posterior probabilities (< 0.95) that $\omega > 1$. Sites 215-219 are close to an indel mutation and the alignment ambiguity in this region could give rise to false signals of positive selection.

5 Discussion

One can learn more about the nature and strength of various physicochemical forces shaping the evolution of positively selected proteins by studying the patterns of nonsynonymous substitutions. Some sites on a positively selected protein may have only undergone neutral nonsynonymous mutations and therefore may not be the targets of natural selection, although they would elevate the estimated nonsynonymous rate. At other sites, there may have been a comparable number of nonsynonymous substitutions but these may have been of a particularly advantageous physicochemical nature (e.g. volume-altering substitutions) thus making them the targets of selection. Both types of sites would have similar posterior probabilities of being under an elevated rate ratio of nonsynonymous to synonymous substitutions ($\omega > 1$). However, the latter type of sites will have higher posterior probabilities of being under volume-altering pressure ($\gamma_v > 1$) than the former. Thus, such sites may be highlighted by an analysis that focuses on the rate of particular nonsynonymous substitutions relative to all other substitution at the codon level. The novelty of this approach stems from allowing both synonymous (structurally-silent and hence functionally-silent) as well as certain tolerable nonsynonymous (not structurally-silent, but functionally-silent) substitutions to specify the scaling rate relative to which the rate of certain functionally-noisy nonsynonymous substitutions are estimated. For some partially understood systems, a user-specified intra-partition scaling rate may allow system-specific hypotheses to be tested. Our method can be thought of as providing a naive alternative to the standard assumption of equivalence between the synonymous rate and the rate of neutral evolution. Our approach also allows for the mining of more information by looking at several codon models induced by different partitions.

One weakness of our approach is the simplifying assumption that no recombination is occuring. Most models of molecular evolution assume that the only source of diversity is point mutations. However, both intra-locus and inter-locus recombination are thought to play a role in the generation of polymorphisms at the MHC (Carrington, 1999). Ignoring recombination partly amounts to compromising for one‘average’ tree for all sites, which is at best a projection of the true ancestral recombination graph residing in a correlated product treespace. Anisimova et al. (2003) reported that while LRTs in a maximum likelihood framework are robust to low levels of recombination, at higher levels the tests may mistake recombination as evidence for positive selection. They find that the M7-M8 likelihood ratio test to detect positive selection is the least affected by recombination. Based on simulations of the Hepatitis D small antigen gene, they also found that the Bayes' prediction of sites under positive selection, based on their posterior probabilities being > 0.95 under the parameter estimates of the M8 model, had a high accuracy of 91% even when there were on average 46.7 recombination events in the history of a sample of 30 sequences. Since SSI is observed to segregate as a single Mendelian trait, the *S*-locus genes are thought to be tightly linked (Casselman et al., 2000). Sequence divergence between *S*-haplotypes, extensive genic rearrangements, and the presence of repetitive elements are factors thought to contribute to the suppression of both intergenic and intragenic recombination at this locus (Boyes and Nasrallah, 1993; Boyes et al., 1997). A recombination analysis of a population segregating for the S6 haplotype, in which the *SRK* and *SCR* genes are separated by more than 200 Kb of sequence (J.B. Nasrallah, unpublished observations), failed to uncover crossovers between these genes in 800 chromosomes analysed (Casselman et al., 2000). This result indicates that recombination in the *S* locus is a rare event. A recent study by Charlesworth et al. (2003) looked at patterns of linkage disequillibrium in SRK alleles from *A. lyrata* and found no evidence supportive of recombination occuring in these genes.

In spite of the possibility for recombination in MHC, conclusions of physicochemical pressures operating on the six class-I MHC proteins, inferred from LRTs as well as the posterior prediction of sites under selective pressure to alter various physicochemical properties, are in agreement with other empirical studies (see previous section). The robustness is partly due to lower levels of recombination in the class-I telomeric region of HLA-A and lack of crossovers in regions between HLA-B and HLA-C, unlike the class-II region with its recombination hotspots (Carrington, 1999). Variation in genomic arrangement that disturbs homology upon pairing is thought to be the cause of reduced recombination. Our very stringent posterior probability cut-off of 0.95 also assures a higher accuracy of detecting positively selected sites (Anisimova et al., 2003). We expect our analysis of the SRK to be at least as robust as that of the MHC due to much lower levels of recombination at the *S* locus. Nonetheless, it is possible for just a few recombination events of the right size and at the right time, to cause distortions to these methods which assume their absence.

As with the class I MHC protein, the selected sites in the SRK protein may be involved in ligand recognition. Therefore, identification of positively selected sites may also help with the identification of regions that play a functional role in female-male (SRK-SCR) interaction and co-evolution. In the absence of a 3D structure, this information may be useful in proposing hypotheses for empirical studies aimed at deciphering the regions responsible for SRK-SCR interaction. Ultimately protein structure and empirical studies are necessary to corroborate the functional significance of the sites identified in this study.

The major weakness of our approach comes from analyzing these partition-based models one at a time. This formulation precludes any formal comparison of models induced by different partitions through their likelihood ratios. Moreover, for all such partition-based models, the synonymous rate conjoined with the property-conserving rate is assumed to be identical for all sites. A model which allows the partitions themselves to vary across sites is more realistic since each site may tolerate changes in different physicochemical properties to different extents. In other words, a more realistic model would allow the nature of neutral evolution itself, and not merely its rate, to be site-specific. Improvement to our models can be made along this front by constructing models that simultaneously allow more than one partition to be parametrized. For instance, such models at the codon level would allow the simultaneous estimation of finite mixtures of volume-altering substitution rate ($\gamma_v \geq 0$), charge-altering substitution rate ($\gamma_c \geq 0$), as well as the nonsynonymous substitution rate ($\omega \geq 0$) relative to the synonymous substitution rate which could be set to 1. The corresponding site-specific posterior probabilities over the positive 3-D orthant with γ_v, γ_c, and ω as its axes would shed more light on rapid evolution under a more flexible model of site-specific variation in the nature as well as rate of neutral evolution. Such models would provide a natural setting to test hypotheses of physicochemical subfunctionalization, subsequent to gene duplication, especially when recast in the lineage as well as site-specific framework of Yang and Nielsen (2002). The suggested improvements to our approach are within the context of computationally fast and efficient finite mixture models that assume independence across sites. Ultimately, such flexible multi-dimensional finite mixture models of neutral evolution that are simultaneously induced by multiple physicochemical partitions should be further extended to allow interactions between sites.

In the current approach, the partition-based models can only be analyzed one at a time. The rate of some property-altering nonsynonymous substitutions that is allowed to vary across sites is scaled relative to the rate of property-conserving nonsynonymous substitutions as well as all synonymous substitutions pooled together homogeneously across sites. The former precludes any direct comparison of different physicochemical pressures while the latter may not be biologically meaningful. On the other hand, the posterior predictions of sites under elevated property-altering to property-conserving rate ratios for several properties are corroborated by empirical findings in MHC. For these reasons we consider these methods to be tools for data exploration more than a complete framework for testing hypotheses regarding positive selection on physicochemical properties.

Acknowledgements R.S. is supported by the Integrative Graduate Education and Research Traineeship from National Science Foundation grant DGE-9870631, National Science Foundation grant DEB-9602229 to Carlos Castillo-Chavez, and National Science Foundation grant DEB-0089487 to Rasmus Nielsen. K.Y. is supported in part by National Institutes of Health grant GM57527 to June Nasrallah. Z.Y. is supported by grants from the Biotechnology and Biological Sciences Research Council (UK) and the Human Frontier Science Program (EU). R.S. thanks Willie Swanson for guiding him through various phylogenetic tools, Wa Yang and Joel Bielewski for the most insightful discussions on physicochemical selection, Stephane Aris-Brosou for discussions on Bayesian confidence sets on tree spaces, Aardra Pontis for discussions on the SSI system, and Akendra De, Matt Dimmic, Richard Durrett, and two anonymous reviewers for comments.

References

Anisimova M, Bielawski JP, Yang Z (2001) Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution Mol Biol Evol 18: 1585–1592

Anisimova M, Nielsen R, Yang Z (2003) Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites Genetics 03: 1229–1236

Barber LD, Percival L, Arnett KL, Gumperz KE, Chen L, Parham P (1997) Polymorphism in the $\alpha 1$ helix of the HLA-B heavy chain can have an overriding influence on peptide-binding specificity Jnl Immunol 158: 1660–1669

Black FL, Hedrick PW (1997) Strong balancing selection at HLA loci: Evidence from segregation in South Amerindian families Proc Natl Acad Sci USA 94: 12452–12456

Boyes DC, Nasrallah JB (1993) Physical linkage of the SLG and SRK genes at the self-incompatibility locus of *Brassica oleracea* Mol Gen Genet 236: 369–373

Boyes DC, Nasrallah ME, Vrebalov J, Nasrallah JB (1997) The self-incompatibility (S) haplotypes of *Brassica* contain highly divergent and rearranged sequences of ancient origin Plant Cell 9: 237–247

Carrington M (1999) Recombination within the human MHC Immunol Rev 167: 245–256

Casselman AL, Vrebalov J, Conner JA, Singhal A, Giovannoni J, Nasrallah ME, Nasrallah JB (2000) Determining the physical limits of the *Brassica* S locus by recombinational analysis Plant Cell 12: 23–33

Charlesworth D, Bartolomé C, Schierup MH, Mable K (2003) Haplotype structure of the stigmatic self-incompatibility gene in natural populations of *Arabidopsis lyrata* Mol Biol Evol 20: 1741–1753

Clarke B (1970) Selective constraints on amino-acid substitutions during the evolution of proteins Nature 228: 159–160

Creighton TE (1996) Proteins: structures and molecular properties W. H. Freeman and co. , New York

Dagan T, Talmor Y, Graur D (2002) Ratios of radical to conservative amino acid replacement are affected by mutational and compositional factors and may not be indicative of positive Darwinian selection Mol Biol Evol 19: 1022–1025

Doherty PC, Zinkernagel RM (1975) Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex Nature 256: 50–52

Domenech N, Santos-Aguado J, Lopez de Castro JA (1991) Antigenicity of HLA-A2 and HLA-B7. loss and gain of serological determinants induced by site-specific mutagenesis at residues 62-80 Hum Immunol 30: 140–146

Dwyer KG, Balent MA, Nasrallah JB, Nasrallah ME (1991) DNA sequences of self-incompatibility genes from *Brassica campestris* and *B. oleracea*: polymorphism predating speciation Plant Mol Biol 16: 481–486

Epstein CJ (1967) Non-randomness in amino-acid changes in the evolution of homologous proteins Nature 215: 355–359

Falk K, Rotzschke O, Stevanovic S, Jung G, H-G R (1991) Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules Nature 351: 290–296

Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach Jnl Mol Evol 17: 368–376

Grantham R (1974) Amino acid difference formula to help explain protein evolution Science 185: 862–864

Guo HC, Madden DR, Silver ML, Jardetzky TS, Gorga JC, Strominger JL, Wiley DC (1993) Comparison of the p2 specificity pocket in the three human histocompatibility antigens: HLA-A*6801, HLA-A*0201, and HLA-B*2705 Proc Natl Acad Sci USA 90: 8053–8057

Hedrick PW, Thomson G (1983) Evidence for balancing selection at HLA Genetics 104: 449–456

Hedrick PW, Whittam TS, Parham P (1991) Heterozygosity at individual amino acid sites: Extremely high levels for HLA-A and -B genes Proc Natl Acad Sci USA 88: 5897–5901

Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection Nature 335: 167–170

Hughes AL, Ota T, Nei M (1990) Positive Darwinian selection promotes charge profile divesity in the antigen-binding cleft of class I major-histocompatibility-complex molecules MBE 7: 515–524

Hulsmeyer M, Hillig RC, Volz A, Ruhl M, Schroder W, Saenger W, Ziegler A, Uchanska-Ziegler B (2002) HLA-B27 subtypes differentially associated with disease exhibit subtle structural alterations Jnl Biol Chem 277: 47844–47853

Kachroo A, Nasrallah ME, Nasrallah JB (2002) Self-incompatibility in the Brassicaceae: receptor-ligand signaling and cell-to-cell communication Plant Cell 14: S227–S238

Kimura M (1983) The neutral theory of molecular evolution Cambridge University Press, Cambridge, England

Macdonald WA, Purcell AW, Mifsud NA, Ely LK, Williams DS, Chang L, Gorman JJ, Clements CS, Kjer-Nielsen L, Koelle DM, Burrows SR, Tait BD, Holdsworth R, Brooks AG, Lovrecz GO, Lu L, Rossjohn J, Mc-

Cluskey J (2003) A naturally selected dimorphism within the HLA-B44 supertype alters class I structure, peptide, repertoire, and T-cell recognition Jnl Exp Med 198: 679–691

Madden DR (1995) The three-dimensional structure of peptide-MHC complexes Annu Rev Immunol 13: 587–622

Markow T, Hedrick PW, Zuerlein K, J D, Martin J, Vyvial T, Armstrong C (1993) HLA polymorphism in the Havasupai: evidence for balancing selection Am Jnl Hum Genet 53: 943–952

Miyata T, Miyazawa S, Yasunaga T (1979) Two types of amino acid substitution in protein evolution Jnl Mol Evol 12: 219–236

Nasrallah JB (2002) Recognition and rejection of self in plant reproduction Science 296: 305–308

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions MBE 3: 418–426

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

Nishio T, Kusaba M (2000) Sequence diversity of SLG and SRK in *Brassica oleracea L.* Ann Botany 85: 141–146

Pakula AA, Sauer RT (1989) Genetic analysis of protein stability and function Annu Rev Genet 23: 289–310

Reche PA, Reinherz EL (2003) Sequence variability analysis of human class I and class II MHC molecules: functional and structural correlates of amino acid polymorphisms Jnl Mol Biol 331: 623–641

Sato K, Nishio T, Kimura R, Kusaba M, Suzuki T, Hatakeyama K, Ockendon DJ, Satta Y (2002) Coevolution of the S-locus genes SRK, SLG and SP11/SCR in *Brassica oleracea* and *B. rapa* Genetics 162: 931–940

Schadt E, Lange K (2002) Codon and rate variation models in molecular phylogeny Mol Biol Evol 19: 1534–1549

Schopfer CR, Nasrallah ME, Nasrallah JB (1999) The male determinant of self-incompatibility in *Brassica* Science 286: 1697–1700

Smith KJ, Reid SW, Harlos K, McMichael AJ, Stuart DI, Bell JI, Jones EY (1996) Bound water structure and polymorphic amino acids act together to allow the binding of different peptides to MHC class I HLA-B53 Immunity 4: 215–228

Sneath PHA (1966) Relations between chemical structure and biological activity Jnl Theor Biol 12: 157–195

Stein JC, Dixit R, Nasrallah ME, Nasrallah JB (1996) SRK, the stigma-specific S locus receptor kinase of *Brassica*, is targeted to the plasma membrane in transgenic tobacco Plant Cell 8: 429–445

Swanson WJ, Yang Z, Wolfner MF, Aquadro CF (2001) Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals Proc Natl Acad Sci USA 98: 2509–2514

Uyenoyama M (1995) A genaralized least-squares estimate for the origin of sporophytic self-incompatibility Genetics 139: 975–992

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

Yang Z, Nielsen R, Goldman N, Pedersen AMK (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites Genetics 155: 431–449

Yang Z, Nielsen R, Hasegawa M (1998) Models of amino acid substitution and applications to mitochondrial protein evolution Mol Biol Evol 15: 1600–1611

Young ACM, Nathenson SG, Sacchettini JC (1995) Structural studies of class I major histocompatibility complex protein: insights into antigen presentation FASEB Jnl 9: 26–36

Zhang J (2000) Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes Jnl Mol Evol 50: 56–68

Zuckerkandl E, Pauling L (1965) Evolutionary divergence and convergence in proteins in: Bryson V, Vogel J (eds.) Evolving genes and proteins Academic press, New York

List of Figures

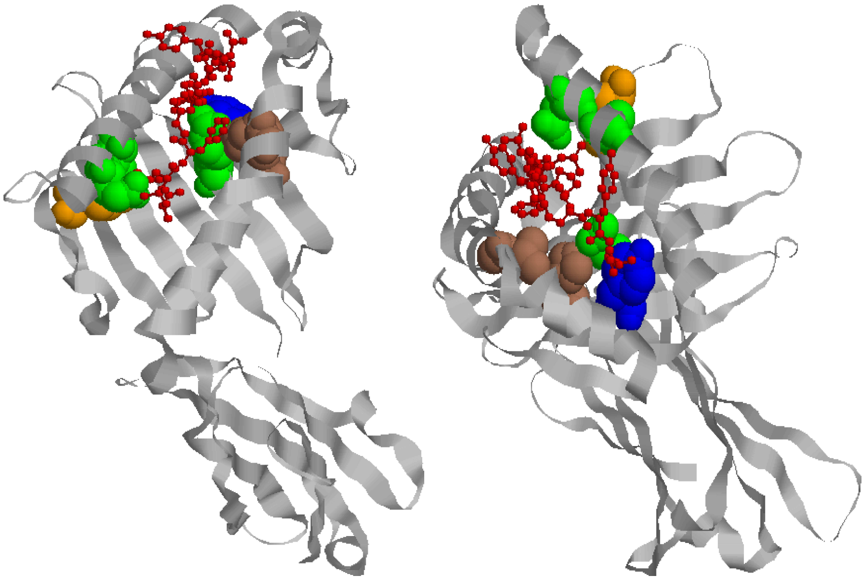

Fig. 1 Sites in the MHC CLASS I protein with high posterior probability ($>$ 0.95) of being under various physicochemical selective pressures. Sites 114 and 156 (both colored brown) are under AA-altering ($\omega > 1$) and charge-altering ($\gamma_c > 1$) pressures. Site 45(orange) is under charge-altering ($\gamma_c > 1$) as well as polarity &/or volume-altering ($\gamma_{pv} > 1$) pressures. Sites 63, 67, and 97 (all green) are under volume-altering ($\gamma_v > 1$) as well as polarity &/or volume-altering ($\gamma_{pv} > 1$) pressures. Finally, site 116 (blue) is under polarity-altering ($\gamma_p > 1$) pressure. The numbering of sites corresponds to the HLA-A2 sequence in the protein data bank file 1QSE. The viral peptide is shown in red.

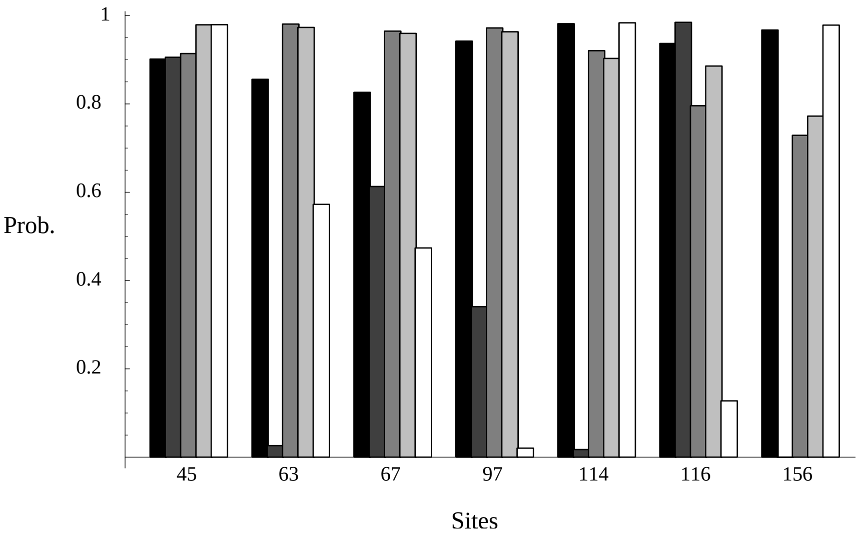

Fig. 2 Posterior probabilities of $\omega > 1$ (black), $\gamma_p > 1$ (dark gray), $\gamma_v > 1$ (gray), $\gamma_{pv} > 1$ (light gray), and $\gamma_c > 1$ (white) at the 7 selected sites in MHC.

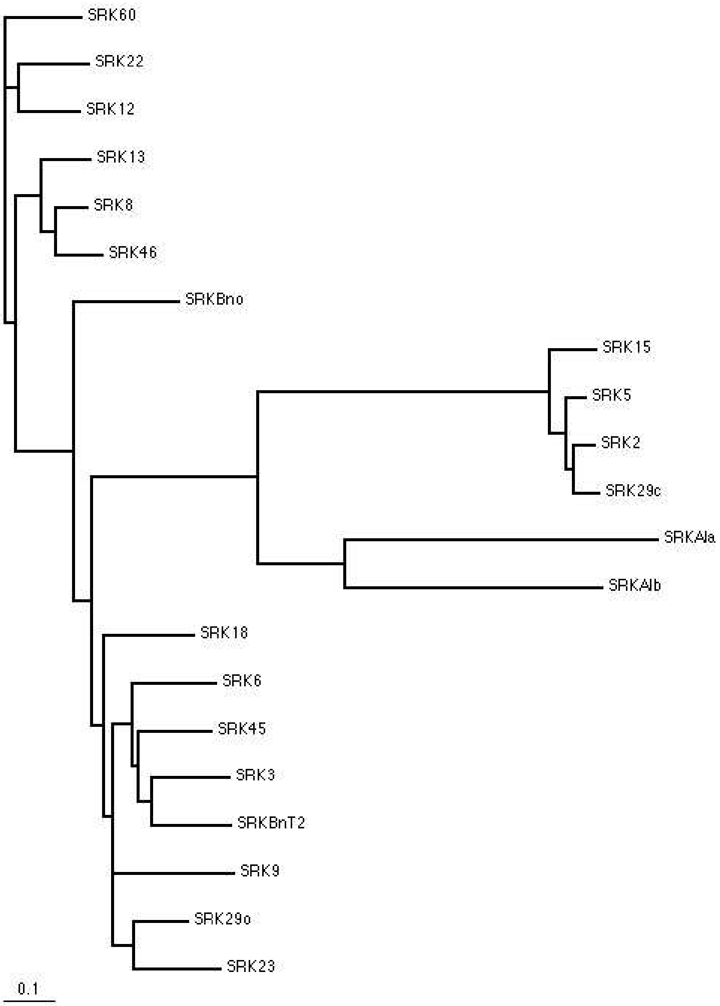

Fig. 3 Maximum likelihood tree under the M8 model for the 21 SRK sequences.

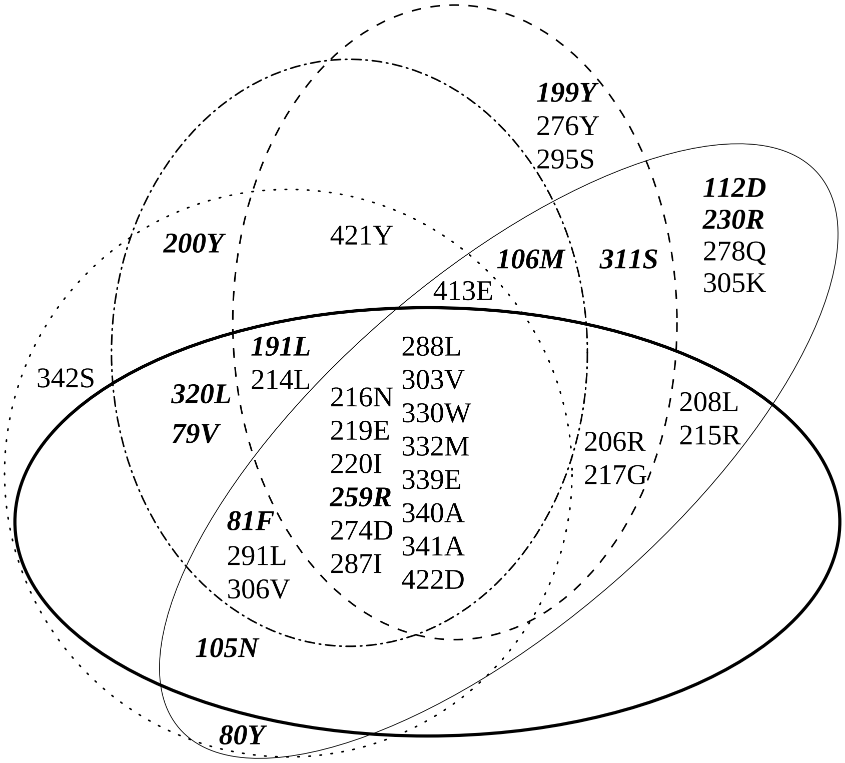

Fig. 4 Sites in SRK with high posterior probability (> 0.95) of being under selective pressure to alter (i) polarity (dashed ellipse), (ii) volume (dotted circle), (iii) polarity and/or volume (dash-dotted ellipse), (iv) charge (thin solid ellipse) and (v) amino acid (thick solid ellipse). The sites shown in bold face italics are outside the HVR and CVR regions. The numbering of sites with their corresponding amino acids is that of *B. oleracea* SRK60.

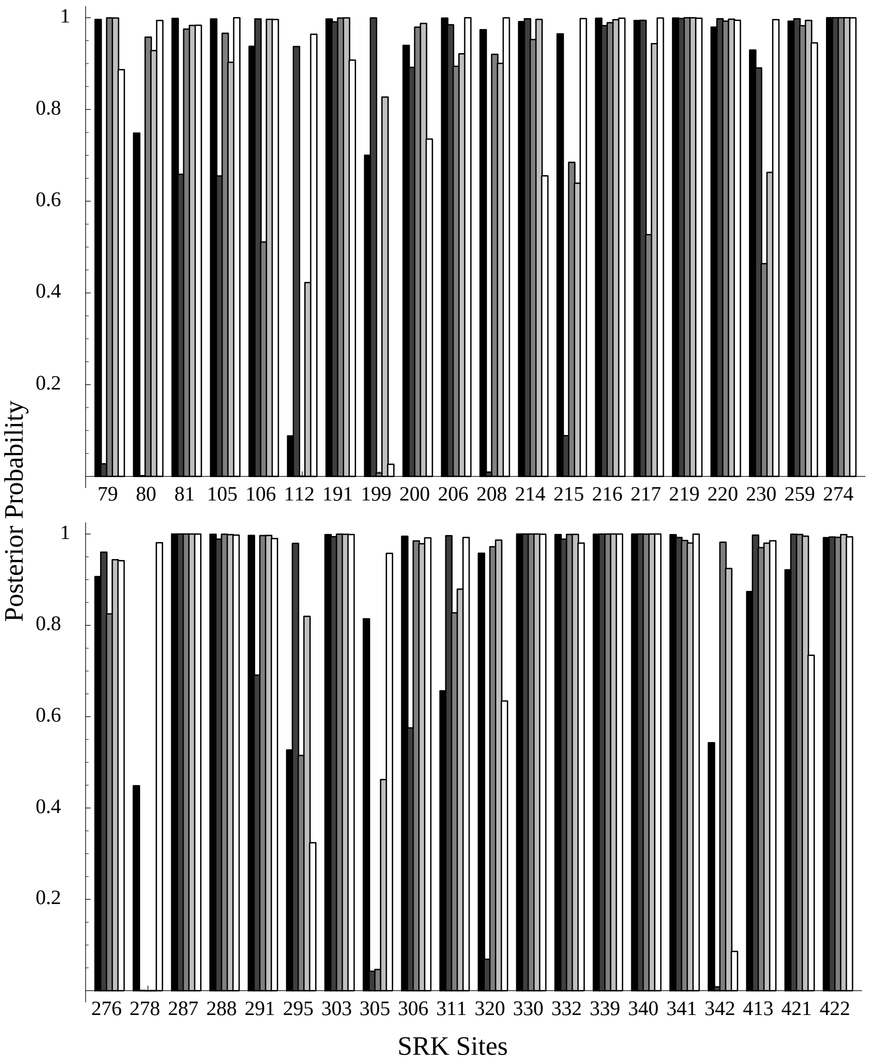

Fig. 5 Posterior probabilities of $\omega > 1$ (black), $\gamma_p > 1$ (dark gray), $\gamma_v > 1$ (gray), $\gamma_{pv} > 1$ (light gray), and $\gamma_c > 1$ (white) at the 40 selected sites in SRK.

List of Tables

Table 1 Partitions

Property $\mathcal{A}$ [$\gamma_{\mathcal{A}}$]	Partition $\mathcal{A}$ of the set of amino acids
Distinctness [ω]	Y, W, H, K, R, E, Q, T, D, N, S, C, M, F, I, V, L, A, G, P
Polarity [γ_p]	polar: {Y, W, H, K, R, E, Q, T, D, N, S, C}, non-polar: {M, F, I, V, L, A, G, P}
Volume [γ_v]	large: {L, I, F, M, Y, W, H, K, R, E, Q}, small: {A, G, C, S, T, D, N, P, V}
Polarity and/or Volume [γ_{pv}]	large polar: {Y, W, H, K, R, E, Q}, small polar: {T, D, N, S, C}, small non-polar: {A, G, P, V}, large non-polar: {L, I, F, M}
Charge [γ_c]	positively charged: {H, K, R}, negatively charged: {D, E}, uncharged: {A, N, C, Q, G, I, L, M, F, P, S, T, W, Y, V}

Table 2 Gene Sequences

Source organism	Gene Name [GenBank Acc. No.]
Homo sapiens	HLA-B*3902 [M94053], HLA-B18 [M24039], HLA-Bw42 [M24034], HLA-A2 SLU [Z27120], HLA-A11E [X13111], HLA-Aw74 [X61701]
Brassica oleracea	SRK6 [M76647], SRK3 [X79432], SRK29o [Z30211], SRK60 [AB032474], SRK18 [AB032473], SRK13 [SEG_AB024419S], SRK23 [AB013720], SRK15 [Y18260], SRK5 [Y18259], SRK2 [AB024416]
Brassica campestris (syn. *B. rapa*)	SRK22 [AB054061], SRK29c [E15797], SRK45 [E15795], SRK46 [SEG_AB013717S], SRK12 [D38564], SRK9 [D30049], SRK8 [D38563]
Brassica napus	SRKBnT2[U00443], SRKBno[M97667]
Arabidopsis lyrata	SRKAla [AB052755], SRKAlb [AB052756]

Table 3 Likelihood Ratio Tests for MHC Class-I HLA

Selective Pressure	Model	ℓ	Parameter estimates	$-2\Delta\ell$	P
nonsynonymous	M7	-2416.94	p=0.011, q=0.022, κ=2.41	14.76	$\ll$0.01
(dn/ds)	M8	-2409.56	p=0.104, q=0.27, κ=2.56		
			p_1=0.073, ω=4.06		
Polarity-altering	M7p	-2442.18	p=0.19, q=0.49, κ=2.42	9.86	$\ll$0.01
	M8p	-2437.25	p=0.47, q=1.46, κ=2.69		
			p_1=0.015, γ_p=44.85		
Volume-altering	M7v	-2439.71	p=0.18, q=0.35, κ=2.41	9.52	$\ll$0.01
	M8v	-2434.95	p=0.41, q=2.15, κ=2.47		
			p_1=0.14, γ_v=2.77		
Polarity &/or	M7pv	-2435.96	p=0.18, q=0.39, κ=2.42	10.74	$\ll$0.01
Volume-altering	M8pv	-2430.59	p=0.46, q=2.14, κ=2.52		
			p_1=0.12, γ_{pv}=3.04		
Charge-altering	M7c	-2439.02	p=0.19, q=0.42, κ=2.54	14.64	$\ll$0.01
	M8c	-2431.70	p=0.48, q=1.69, κ=2.68		
			p_1=0.06, γ_c=6.45		

Table 4 Likelihood Ratio Tests for SRK

Selective Pressure	Model	ℓ	Parameter estimates	$-2\Delta\ell$	P
nonsynonymous (dn/ds)	M7 M8	-17627.95 -17507.17	p=0.30, q=0.48, κ=2.03 p=0.36, q=0.62, κ=2.21 p_1=0.059, ω=3.28	241.6	$\ll$0.01
Polarity-altering	M7p M8p	-17955.09 -17858.48	p=0.18, q=0.41, κ=2.04 p=0.44, q=1.25, κ=2.16 p_1=0.065, γ_p=5.00	193.2	$\ll$0.01
Volume-altering	M7v M8v	-17873.72 -17760.51	p=0.24, q=0.44, κ=1.99 p=0.45, q=0.97, κ=2.11 p_1=0.054, γ_v=4.68	226.4	$\ll$0.01
Polarity &/or Volume-altering	M7pv M8pv	-17771.62 -17656.94	p=0.25, q=0.46, κ=1.98 p=0.47, q=1.06, κ=2.12 p_1=0.068, γ_{pv}=3.58	229.4	$\ll$0.01
Charge-altering	M7c M8c	-17917.87 -17788.99	p=0.22, q=0.37, κ=2.07 p=0.40, q=0.83, κ=2.21 p_1=0.07, γ_c=5.12	257.76	$\ll$0.01