Published in final edited form as:
Immunogenetics. 2013 September ; 65(9): 655–665. doi:10.1007/s00251-013-0714-9.

MHCcluster, a method for functional clustering of MHC molecules

Martin Thomsen,
Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Building 208, Kemitorvet, Lyngby 2800, Denmark

Claus Lundegaard,
Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Building 208, Kemitorvet, Lyngby 2800, Denmark. ALK, Bøge Allé 6, 2970 Hørsholm, Denmark

Søren Buus,
Laboratory of Experimental Immunology, University of Copenhagen, Copenhagen N, Denmark

Ole Lund, and
Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Building 208, Kemitorvet, Lyngby 2800, Denmark

Morten Nielsen
Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Building 208, Kemitorvet, Lyngby 2800, Denmark. Instituto de Investigaciones Biotecnológicas, Universidad Nacional de San Martín, San Martín, Buenos Aires, Argentina

Morten Nielsen: mniel@cbs.dtu.dk

Abstract

The identification of peptides binding to major histocompatibility complexes (MHC) is a critical step in the understanding of T cell immune responses. The human MHC genomic region (HLA) is extremely polymorphic comprising several thousand alleles, many encoding a distinct molecule. The potentially unique specificities remain experimentally uncharacterized for the vast majority of HLA molecules. Likewise, for nonhuman species, only a minor fraction of the known MHC molecules have been characterized. Here, we describe a tool, *MHCcluster*, to functionally cluster MHC molecules based on their predicted binding specificity. The method has a flexible web interface that allows the user to include any MHC of interest in the analysis. The output consists of a static heat map and graphical tree-based visualizations of the functional relationship between MHC variants and a dynamic TreeViewer interface where both the functional relationship and the individual binding specificities of MHC molecules are visualized. We demonstrate that conventional sequence-based clustering will fail to identify the functional relationship between molecules, when applied to MHC system, and only through the use of the predicted binding specificity can a correct clustering be found. Clustering of prevalent HLA-A and HLA-B alleles using *MHCcluster* confirms the presence of 12 major specificity groups (supertypes) some however with highly divergent specificities. Importantly, some HLA molecules are shown not to fit any supertype classification. Also, we use *MHCcluster* to show that chimpanzee MHC class I molecules have a reduced functional diversity compared to that of HLA class I molecules. *MHCcluster* is available at www.cbs.dtu.dk/services/MHCcluster-2.0.

Correspondence to: Morten Nielsen, mniel@cbs.dtu.dk.

Keywords

MHC; HLA; Binding motif; Functional clustering; MHC specificity; Supertypes

Introduction

Major histocompatibility complex (MHC) molecules play a central role in generating specific T cell-mediated immune responses. T cells scrutinize small peptide fragments, also called epitopes, presented in a complex with MHCs on the surface of most cells in the host. Cytotoxic T cells kill cells that present peptides of foreign or abnormal origin in a complex with MHC class I molecules. T helper cells, on the other hand, orchestrate the immune response by stimulating other immune cells and are stimulated by antigen-presenting cells that display peptides in complex with MHC class II molecules. The binding of peptides to MHC molecules is hence a prerequisite for T cell immunogenicity. Identifying which peptides will be presented in complex with a given MHC molecule is therefore of pivotal importance for our understanding of cellular immunity.

In general, MHC molecules are highly specific, binding only a minor fraction of the set of possible peptides (Yewdell and Bennink 1999; Rao et al. 2009). Moreover, the genomic region encoding MHC molecules is extremely polymorphic comprising several thousand alleles, many encoding a distinct molecule, making the peptide-binding repertoire of each individual unique. The most recent IMGT/human leukocyte antigen (HLA) database (Robinson and Marsh 2007) contains close to 5,000 HLA (the human version of MHC) class I protein sequences. This immense polymorphism of MHC molecules makes it a very costly endeavor to experimentally characterize the binding specificity of each molecule. Despite the significant experimental progress in high-throughput screening technologies (Harndahl et al. 2009, 2011), less that 80 HLA class I molecules have to this day been characterized with peptide binding data, allowing an accurate characterization of their binding motif (data taken from the IEDB; Vita et al. 2010). For nonhuman species including life-stock animals, the situation is even worse. Here, only a minor fraction of the known MHC molecules have been functionally characterized.

Due to the high selectivity of the MHC molecules, major efforts have been dedicated to characterize their binding specificity and several in silico methods have been developed allowing prediction of the binding affinity of peptides to MHC molecules (reviewed in Lundegaard et al. (2010) and Nielsen et al. (2010b)). These state-of-the-art methods make it possible to predict not only the MHC binding repertoire of any MHC molecule of interest (Hoof et al. 2009; Nielsen et al. 2010a; Karosiene et al. 2011), but also to characterize the subtle difference in the MHC specificities imposed by the allelic difference (Erup Larsen et al. 2011).

Not all MHC molecules are equally different in term of function, and several approaches have been described aiming to perform clustering of MHC molecules based on different measures of (functional) similarity (Sette and Sidney 1999; Doytchinova et al. 2004; Lund et al. 2004; Hertz and Yanover 2007). In 1999, Sette and Sidney proposed the HLA class I supertype concept, proposing that allelic variants within a supertype would share a large functional overlap, and nine such supertypes could cover the HLA class I functional space (Sette and Sidney 1999). Using data of known HLA class I ligands, Lund et al. refined this in 2004, and suggested the presence of three additional functional clusters (Lund et al. 2004). A limiting factor for the HLA clustering approach suggested by Lund et al. is the need for known ligands for the MHC molecules interest. In the original *NetMHCpan* publication (Nielsen et al. 2007), we therefore suggested the use of correlations between

predicted binding affinities to perform functional clustering of HLA molecules and demonstrated that this approach could accurately reproduce the earlier proposed 12 HLA supertypes (and similar results have been shown for MHC class II; Nielsen et al. 2008). The functional clustering proposed by *NetMHCpan* demonstrated that many HLA molecules are characterized by specificities that are poorly characterized by the common 12 supertypes. This underlines an important shortcoming of the supertype concept.

Here, we describe a freely available web server, *MHCcluster*, implementing the functional clustering procedure described above to functionally cluster MHC molecules based on their predicted binding specificity. The method can be applied for both MHC classes I and II molecules for any MHC molecule with a known protein sequence covered by the *NetMHCpan* and *NetMHCIIpan* prediction methods (that is any MHC class I molecule and any HLA-DR class II molecule). The method has a flexible web interface that allows the user to include any MHC of interest in the analysis. The output from *MHCcluster* consists of a static heat map and graphical tree-based visualizations of the functional relationship between MHC variants and a dynamic TreeViewer interface where both the functional relationship and the individual binding specificities of MHC molecules are visualized.

We illustrate the power of the *MHCcluster* method in three distinct settings. First, we compare conventional sequence-based clustering to the functional clustering of *MHCcluster* and demonstrate situations where a sequence-based clustering, in contrast to *MHCcluster*, fails to identify the correct functional relationship between alleles. Next, we apply *MHCcluster* to the HLA-A and HLA.B. system investigating to what extent the common 12 HLA supertypes give an accurate representation of the functional diversity. Lastly, we use the method to confirm earlier findings (van Deutekom et al. 2011) demonstrating that chimpanzee MHC class I molecules have a reduced functional diversity compared to that of HLA class I molecules.

Materials and methods

Method

The *MHCcluster* server allows the user to select a set of MHC alleles of interest including the option of uploading a set of full-length MHC I protein sequences and the server returns an unrooted tree and a heat map visualizing the functional similarities between the MHC molecules. The vehicles underlying the *MHCcluster* server are the *NetMHCpan* (version 2.7) *and NetMHCIIpan* (version 2.1) prediction methods. For each selected MHC allele, the *MHCcluster* method predicts its binding to a set of predefined natural peptides. Next, the similarity between any two MHC molecules is estimated from the correlation between the predictions of the union of the top 10 % strongest binding peptides for each allele (the threshold value can be altered by the user). This similarity is 1 if the two molecules have a perfect binding specificity overlap and −1 if the two molecules share no specificity overlap. Given this similarity, a distance between two molecules is defined as 1–similarity. The distance matrix is converted to an UPGMA (unweighted pair group method with arithmetic mean distance tree. To estimate the significance of the MHC distance) tree, a large set of distance trees is generated using the bootstrap method and a final tree is summarized in the form of a “greedy” consensus tree with corresponding branch bootstrap values.

Sequence logos

As part of the new *MHCcluster*, a sequence logo for each allele is generated using the *Seq2Logo* service (Thomsen and Nielsen 2012). The logos are created from the top 1 % strongest binding peptides. For MHCII alleles, the logo is constructed from the predicted 9mer binding cores. The sequences used in the logos are clustered using the *Hobohm* 1

algorithm (Hobohm et al. 1992) using a similarity threshold of 63 % to remove redundancy, and pseudo counts are applied with a weight on prior of 200 (Altschul et al. 1997).

Prevalent HLA molecules

Prevalent HLA-A, B, and C molecules were identified for the European population from the dbMHC (NCBI Resource Coordinators 2013) using an allele frequency threshold of 0.5 %. The set of alleles defined as "HLA Prevalent and Characterized" consists of the HLA molecules characterized with more than 50 peptide binding data points and more than 0.5 % worldwide prevalence (as defined by the Allele Frequency Net database (Middleton et al. 2003), for populations characterized with more than 500 fully typed samples).

The MHCcluster server

The submission interface to the *MHCcluster* server is shown in Fig. 1. Here, the users can specify whether they wish to analyze MHC class I or MHC class II molecules, subsequently select the set of molecules to compare (including the option to analyze novel MHC molecules), define how many bootstrap samples to use, the number of peptides to include in the functional correlation analysis, and the threshold used to select peptide from the correlation analysis. To aid the selection of predefined sets of alleles, a "Select All" option is included.

The output from the server consists of a static heat map and graphical tree-based visualizations of the functional relationship between MHC variants and a dynamic TreeViewer interface where both the functional relationship and the individual binding specificities of MHC molecules are visualized. An output example of the static output from the server is shown in Fig. 2. Several result files are available for download including the distance file used to construct the heat map and tree, a zip file with the sequence logos of the predicted binding motif for each MHC molecule analyzed, and a file containing the estimated accuracy of the predicted sequence motifs. This accuracy is estimated from the distance to the nearest MHC molecules included in the training of the peptide binding prediction method (for details, see Hoof et al. (2009)). A value greater than 0.90 indicates high accuracy and values down to 0.70 are considered accurate predictions (Karosiene et al. 2011). Since *MHCcluster* provides prediction of binding motifs for the MHC molecules analyzed, it is important to take the accuracy values into account when interpreting the results.

By clicking on the tree or following the link below the tree, the user is taken to the dynamic TreeViewer interface (see Fig. 3). Here, the user can customize the tree of the functional relationship of the selected MHC molecules in different ways including the option of adding visualization of the individual binding specificities in terms of sequence logos.

The allele liston the right side of the tree lists all the alleles on the tree and presents the user with several functions. Firstly, it allows the user to locate the allele on the tree by showing the motif next to the node on the tree when the mouse cursor hovers over the allele name. Secondly, it allows the user to permanently add the motifs of selected alleles by selecting them on the list. These can be removed again by double clicking on either the motif or the allele, and thirdly, it shows the user which alleles have their motifs shown on the tree. The estimated accuracy value of the predicted sequence motif, a number between 0 and 1, is also provided next to the allele names.

The comparison bar below the tree shows the motifs of the selected alleles side by side to make it fast and easy for the user to compare the motifs of the alleles. The alleles can be

selected in two ways either by selecting alleles on the allele list or by right clicking on the alleles on the tree and choosing a slot in the pop-up menu.

The tree (see Fig. 4) is generated through a simple algorithm, which computes the tree from a Newick file where the branches are spatially distributed according to their relative sizes. After the tree has been computed, it is rescaled to fit the predefined box. The tree is drawn as sticks, and in the terminal of all branches a circle is drawn to represent the node. The corresponding label (if any) is drawn in extension hereof. On the branches with bootstrap values, the value is drawn in a centered position above the branch.

User interaction

After everything is drawn, a few user interaction events are added to the tree elements. The nodes and labels receive a hover event, which shows a motif for the corresponding MHC allele next to the cursor, and they also receive a right-click event, which activates a menu where the user can add the corresponding MHC allele motif to a slot in the comparison bar. In addition, the labels and the permanent motifs (motifs from the selected alleles in the allele list) can now be dragged to any location in the box by left clicking, holding and dragging the element with the cursor. When the user left clicks on a motif or a label, the motif/label is brought to the front of the screen. This feature makes it possible for the user to arrange the overlapping images and labels as preferred. Finally, a copy of TreeViewer, including the tree file and corresponding sequence logos, can be downloaded as a zip file allowing the user to work and generate figures locally.

Results

To illustrate the important difference between sequence and function-based clustering, we generate a clustering of the HLA-A*68:01 and HLA-A*68:02 molecules with a set of HLA-A and HLA-B alleles representing the 12 supertypes using the conventional phylogenetic approach based on the full-length protein sequence of the different HLA molecules. The result of this clustering is shown in Fig. 5.

One thing that is important to note in this figure is the very close distance between the HLA-A*68:01 and HLA-A*68:02 molecules. On the sequence level, these two molecules share close to 99 % similarity differing by only five amino acid substitutions. However, when we look at the binding specificity as represented by the logos, it is apparent that these two molecules are very different in terms of function. The HLA-A*68:01 molecule has an A3 supertype specificity with a preference for basic amino acids at the C terminal whereas HLA-A*68:02 has a mixed A2/A26 specificity matching A26 at the N terminal and A2 at the C terminal.

If we repeat the analysis using the functional clustering of the *MHCcluster* method, we obtain the result displayed in Fig. 6. Using functional clustering, the functional difference between the HLA-A*68:01 and HLA-A*68:02 molecules becomes apparent. In the functional tree, HLA-A*68:01 is shown to have an A3 like specificity whereas HLA-A*68:02 has a mixed A2 and A26 specificity (see heat map in right panel of Fig. 6).

We next apply the functional clustering to a set of 42 prevalent HLA-A and HLA-B molecules. From the results of this analysis (Fig. 7), it is clear that the clustering to a very high degree reproduces the 12 HLA supertypes. However, it is also apparent from the figure that some HLA-A and B molecules are not well characterized by the specificities of the common 12 supertypes. In particular, is it clear that the A3, B7, and B44 “clusters” consist of HLA molecules with highly divergent specificities, and that some molecules like HLA-A*30:01, HLA-A*29:02, and HLA-B*38:01 are poorly characterized by the specificities of

the 12 supertypes. Note, that these molecules all have a predicted accuracy value of 1.00, indicating that the molecules are characterized by the peptide binding data. The calculated binding motifs thus with a high likelihood give a correct representation of the specificity of the molecules in question.

Having demonstrated that the automated *MHCcluster* method is capable of producing a functional clustering of the HLA-A and HLA-B molecules that fits the picture of 12 distinct specificity groups, we apply the method to illustrate how the method in a highly intuitive manner can be used to address questions related to comparative functiomics of MHC genes. This we do by comparing the human HLA-A, -B and chimpanzee Patr-A, -B molecules investigating whether a loss in functional diversity could be observed in the Patr-A loci as suggested by the selective sweep theory (de Groot et al. 2008; van Deutekom et al. 2011). The result of this analysis is given in Fig. 8.

The figure shows a specificity-tree of the 42 prevalent HLA-A and B alleles included in Fig. 7 combined with a set of 31 Patr-A alleles (including the additional chimpanzee Alike MHC class I molecule, Patr-AL) and 47 Patr-B alleles. Note, that more than 92 % of the Patr (and all of the HLA) motifs are predicted with an accuracy value greater than 0.7. The figure supports the notion (de Groot et al. 2008; van Deutekom et al. 2011) that chimpanzees have a reduced MHC class I binding specificity repertoire compared to humans. In particular, the figure suggests that the Patr-A loci has a reduced specificity repertoire compared to the HLA-A loci, and that chimpanzees seem to lack specificities matching the human HLA-A26, and to some extent the HLA-A2 supertypes. We can quantify this observation calculating the pairwise functional distances between molecules within the HLA-A, HLA-B, Patr-A, and Patr-B loci (see Fig. 9). Doing this clearly reveals the great loss of specificity diversity of the Patr-A locus compared to that of the human HLA-A locus. Close to 95 % of the total number of large (>0.8) intralocus functional distances stem from the HLA-A loci, the last 5 % stem from the Patr-A, and none stem from the HLA-B and Patr-B loci. Likewise, 79 % of the pair-wise HLA-A distances are greater than 0.6, whereas less than 45 % of the pairwise Patr-A distances are greater than 0.6. In terms of statistics, these differences are highly significant ($p<0.001$, comparison of ratio). With regards to the two B loci, only minor differences are observed. A few HLA-B specificities seem to be missing from the Patr-B loci (HLA-B08, HLA-B27, and HLA-B39), but in terms of functional diversity they are to some degree compensated for by a set of novel specificities unique to Patr (represented by for instance Patr-B0601, Patr-B1701, and Patr-B2202).

Discussion and conclusion

Functional clustering of MHC molecules is a highly challenging task due to the vast polymorphism of the MHC genomic region and the very delicate relationship between subtle amino acid substitutions and dramatic variations in binding specificity. Here, we have illustrated how conventional sequence-based methods due to this subtle relationship in many cases will fail to produce a correct clustering and functional annotation for MHC molecules.

Given this observation, we argue that clustering and functional annotation for MHC molecules must be made based on information reflecting the peptide binding preference for each molecule and propose the *MHCcluster* method as an effective visual tool to compare functional similarities between large sets of MHC molecules.

The *MHCcluster* method estimates the functional relationship between two molecules from the overlap in prediction binding specificity, and returns a heat map and graphical tree-based visualizations of the functional relationship between MHC variants. Besides these conventional representations of the functional map of the MHC molecules of interest, the *MHCcluster* method provides a dynamic TreeViewer interface where both the functional

relationship and the individual binding specificities of MHC molecules are visualized (the later in terms of sequence logos). This TreeViewer is a unique feature of the *MHCcluster* server that allows in a highly intuitive manner for functional interpretations of the MHC map proposed by the *MHCcluster* method. Earlier methods have been proposed for functional clustering of MHC molecules (Sette and Sidney 1999; Doytchinova et al. 2004; Lund et al. 2004; Hertz and Yanover 2007), and for the browsing of predicted binding motifs of MHC molecules (Rapin et al. 2008; Rapin et al. 2010). But to the best of our knowledge, no method has combined these two approaches allowing for the direct functional mapping of MHC molecules in terms of both clustering and visualization of binding motifs.

Using the *MHCcluster* method, we confirm the existence of the 12 HLA supertypes earlier proposed to characterize the specificity space of HLA-A and HLA-B molecules. However, the analysis also clearly revealed that not all HLA molecules fit equally well into a supertype classification scheme, and that some supertype “clusters” consist of molecules with highly divergent specificities. Finally, moving to nonhuman primates, we compare the MHC class I specificity space of human and chimpanzee using the *MHCcluster* method and demonstrate that the Patr A locus has significantly reduced functional diversity compared to the human HLA-A locus manifested by the almost complete loss of HLA-A2 and HLA-A26 supertype specificities.

In this work, we have focused on demonstrating the use of the *MHCcluster* method to analyze functional diversities of MHC class I molecules. The method is equally well suited for making functional analysis and clustering for MHC class II molecules, and the server does include an option to analyze MHC class II molecules. However, as no pan-specific prediction algorithm currently exists to allow for the prediction of peptide binding to any MHC class II molecule, the analysis is limited to the HLA-DR loci molecules covered by the *NetMHCIIpan* method (Nielsen et al. 2008, 2010a).

In conclusion, we have demonstrated that the *MHCcluster* method can be used as an effective visual tool to compare functional similarities between MHC molecules. The method is highly flexible and allows the user to analyze any MHC variant of interest. *MHCcluster* is available at www.cbs.dtu.dk/services/MHCcluster-2.0.

Even though we here have limited the applications of the *MHCcluster* method to the comparison of functional similarities between large sets of MHC molecule, many other types of important questions that can be addressed by the method. Some could be as a guide to help researchers interpret immunological phenotypic similarities between patients using information about HLA types (i.e., understand for instance why patients with no overlap in HLA types can share an overlap in T cell epitopes), as a guide to see where a specific allele (maybe present at a high frequency in a particular cohort) fits in to the specificity space covered by the common MHCs.

Acknowledgments

MN is researcher at the Argentinean National Research Council (CONICET). This work was supported by NIH grant HHSN272200900045C.

References

Altschul SF, Madden TL, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997; 25:3389–3402. [PubMed: 9254694]

de Groot NG, Heijmans CM, et al. Pinpointing a selective sweep to the chimpanzee MHC class I region by comparative genomics. Mol Ecol. 2008; 17(8):2074–2088. [PubMed: 18346126]

Doytchinova IA, Guan P, et al. Identifiying human MHC supertypes using bioinformatic methods. J Immunol. 2004; 172(7):4314–4323. [PubMed: 15034046]

Erup Larsen M, Kloverpris H, et al. HLArestrictor-a tool for patient-specific predictions of HLA restriction elements and optimal epitopes within peptides. Immunogenetics. 2011; 63(1):43–55. [PubMed: 21079948]

Harndahl M, Justesen S, et al. Peptide binding to HLA class I molecules: homogenous, high-throughput screening, and affinity assays. J Biomol Screen. 2009; 14(2):173–180. [PubMed: 19196700]

Harndahl M, Rasmussen M, et al. Real-time, high-throughput measurements of peptide-MHC-I dissociation using a scintillation proximity assay. J Immunol Methods. 2011; 374(1–2):5–12. [PubMed: 21044632]

Hertz T, Yanover C. Identifying HLA supertypes by learning distance functions. Bioinformatics. 2007; 23(2):e148–e155. [PubMed: 17237084]

Hobohm U, Scharf M, et al. Selection of representative protein data sets. Protein Sci. 1992; 1:409–417. [PubMed: 1304348]

Hoof I, Peters B, et al. NetMHCpan, a method for MHC class I binding prediction beyond humans. Immunogenetics. 2009; 61(1):1–13. [PubMed: 19002680]

Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 2006; 23(2):254–267. [PubMed: 16221896]

Karosiene E, Lundegaard C, et al. NetMHCcons: a consensus method for the major histocompatibility complex class I predictions. Immunogenetics. 2011; 64(3):177–186. [PubMed: 22009319]

Larkin MA, Blackshields G, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007; 23(21): 2947–2948. [PubMed: 17846036]

Lund O, Nielsen M, et al. Definition of supertypes for HLA molecules using clustering of specificity matrices. Immunogenetics. 2004; 55(12):797–810. [PubMed: 14963618]

Lundegaard C, Lund O, et al. Major histocompatibility complex class I binding predictions as a tool in epitope discovery. Immunology. 2010; 130(3):309–318. [PubMed: 20518827]

Middleton D, Menchaca L, et al. New allele frequency database. Tissue Antigens. 2003; 61(5):403–407. http://www.allelefrequencies.net. [PubMed: 12753660]

NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2013; 41(Database issue):D8–D20. [PubMed: 23193264]

Nielsen M, Lundegaard C, et al. NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. PLoS One. 2007; 2(8):e796. [PubMed: 17726526]

Nielsen M, Lundegaard C, et al. Quantitative predictions of peptide binding to any HLA-DR molecule of known sequence: NetMHCIIpan. PLoS Comput Biol. 2008; 4(7):e1000107. [PubMed: 18604266]

Nielsen M, Justesen S, et al. NetMHCIIpan-2.0—improved pan-specific HLA-DR predictions using a novel concurrent alignment and weight optimization training procedure. Immunol Res. 2010a; 6:9.

Nielsen M, Lund O, et al. MHC class II epitope predictive algorithms. Immunology. 2010b; 130(3): 319–328. [PubMed: 20408898]

Rao X, Costa AI, et al. A comparative study of HLA binding affinity and ligand diversity: implications for generating immunodominant CD8+ T cell responses. J Immunol. 2009; 182(3):1526–1532. [PubMed: 19155500]

Rapin N, Hoof I, et al. MHC motif viewer. Immunogenetics. 2008; 60(12):759–765. [PubMed: 18766337]

Rapin N, Hoof I, et al. The MHC motif viewer: a visualization tool for MHC binding motifs. Current protocols in immunology. 2010; 18(18):17.10.1002/0471142735.im1817s88

Robinson J, Marsh SG. The IMGT/HLA database. Methods Mol Biol. 2007; 409:43–60. [PubMed: 18449991]

Sette A, Sidney J. Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. Immunogenetics. 1999; 50:201–212. [PubMed: 10602880]

Thomsen MC, Nielsen M. Seq2Logo: a method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion. Nucleic Acids Res. 2012; 40 (Web Server issue):W281–287. [PubMed: 22638583]

van Deutekom HW, Hoof I, et al. A comparative analysis of viral peptides presented by contemporary human and chimpanzee MHC class I molecules. J Immunology. 2011; 187(11):5995–5600. [PubMed: 22043011]

Vita R, Zarebski L, et al. The immune epitope database 2.0. Nucleic Acids Res. 2010; 38 (database issue):D854–862. [PubMed: 19906713]

Yewdell JW, Bennink JR. Immunodominance in major histo-compatibility complex class I-restricted T lymphocyte responses. Annu Rev Immunol. 1999; 17:51–88. [PubMed: 10358753]

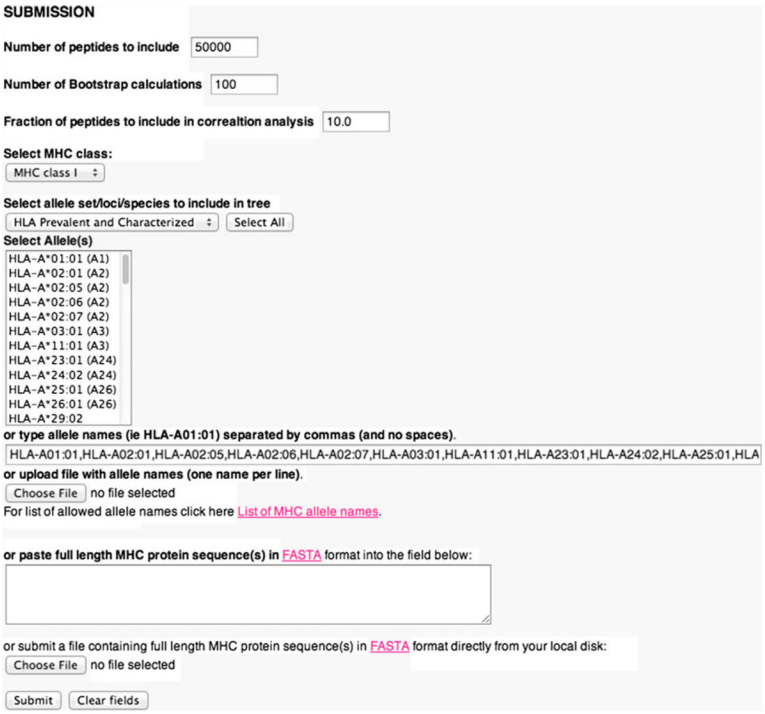

Fig. 1.
Interface to the *MHCcluster* server

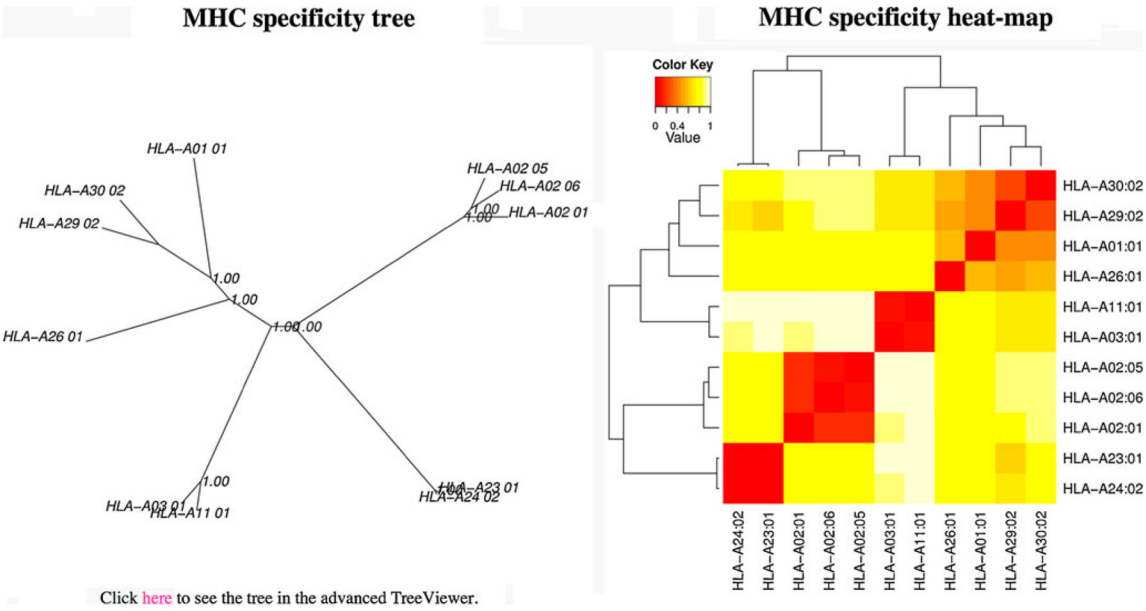

Fig. 2.

The static output from the *MHCcluster* server. The figure displays the functional clustering of 12 HLA-A class I molecules. The *left panel* shows the unrooted tree visualization and the *right panel* shows the corresponding heat map

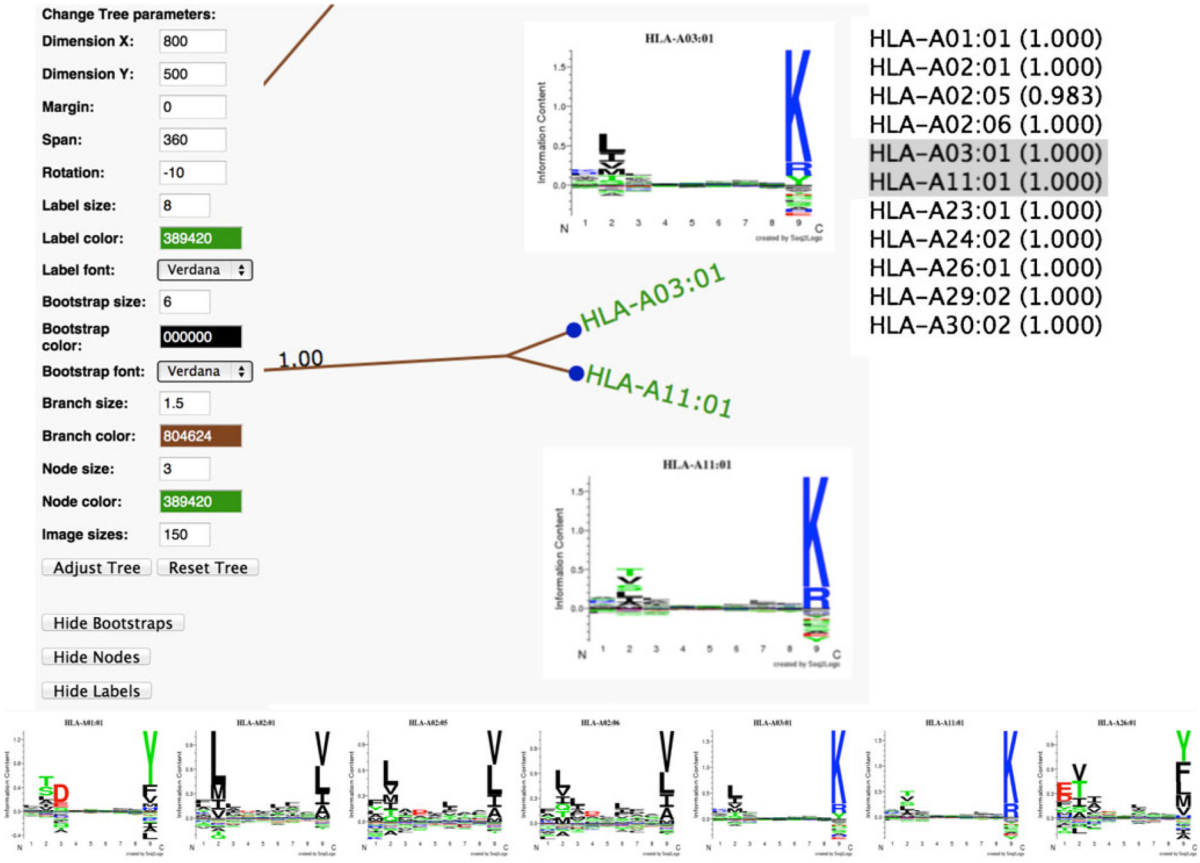

Fig. 3.

The dynamic TreeViewer interface. *Upper left panel* The setting panel of the TreeViewer presents the user with several options to customize the tree. At the top are listed the settings, which directly impact the form and size of the tree. Next, follows the labels and bootstrap settings to adjust the size, color and font of the text in the tree. Thirdly, the branch, node, and image appearance settings are given. Another option is the buttons to update the tree with your changes. Lastly, the show and hide buttons allows to hide bootstrap values, nodes, and labels on the tree. To pick a new color, click on the colored field, then choose the preferred color by moving the mouse cursor over the color palette and then left click to select the color. Hit enter to accept. *Upper right panel* The allele list of the TreeViewer, where the alleles HLA-A*03:01 and HLA-A*11:01 have been selected on the list, and its motif is shown next to the node. *The lower panel* The comparison bar shows the motifs of the selected alleles side by side to make it fast and easy for the user to compare the motifs of the alleles

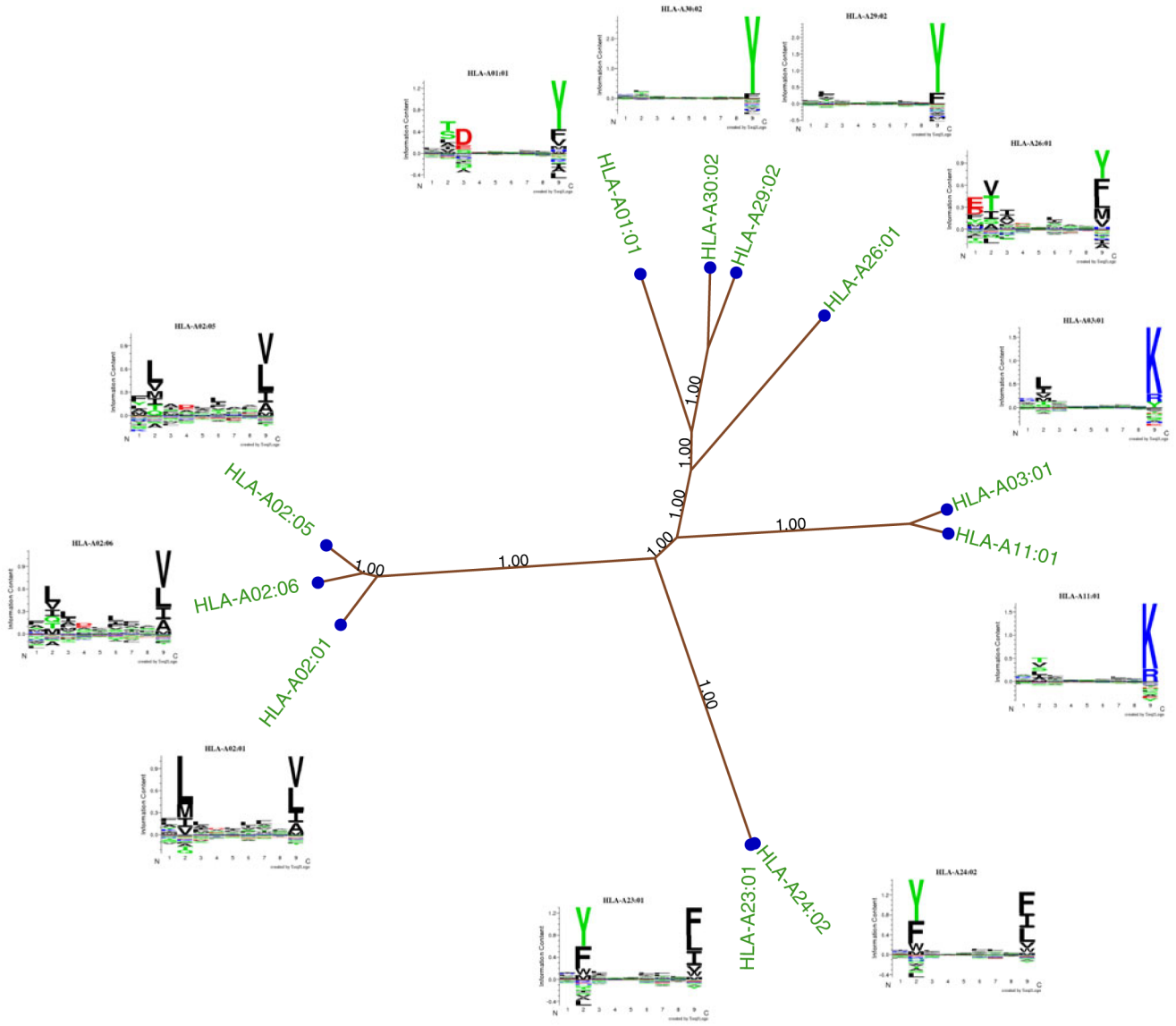

Fig. 4.

The tree is a visualization of the Newick file generated during the *MHCcluster* computations. This tree has been recolored and the motifs of all the alleles have been arranged around the tree to provide the viewer with quick information of the different binding motifs

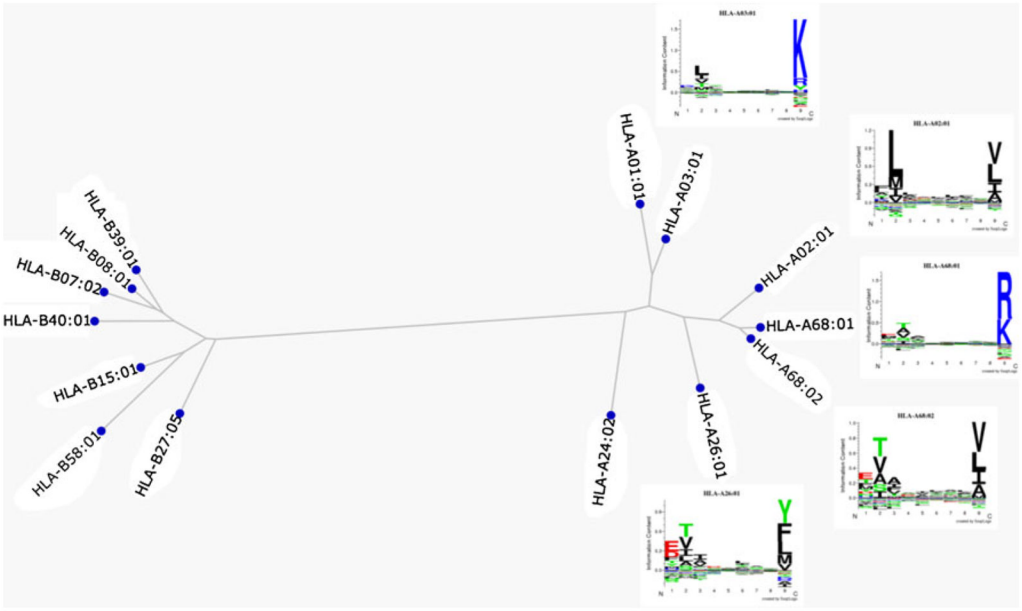

Fig. 5.

Phylogenetic clustering using full-length protein sequences of HLA-A*68:01, HLA-A*68:02 and alleles representing the common 12 HLA supertypes. The clustering was made using ClustalX (Larkin, Blackshields et al. 2007) and the tree was visualized using the TreeViewer of *MHCcluster*

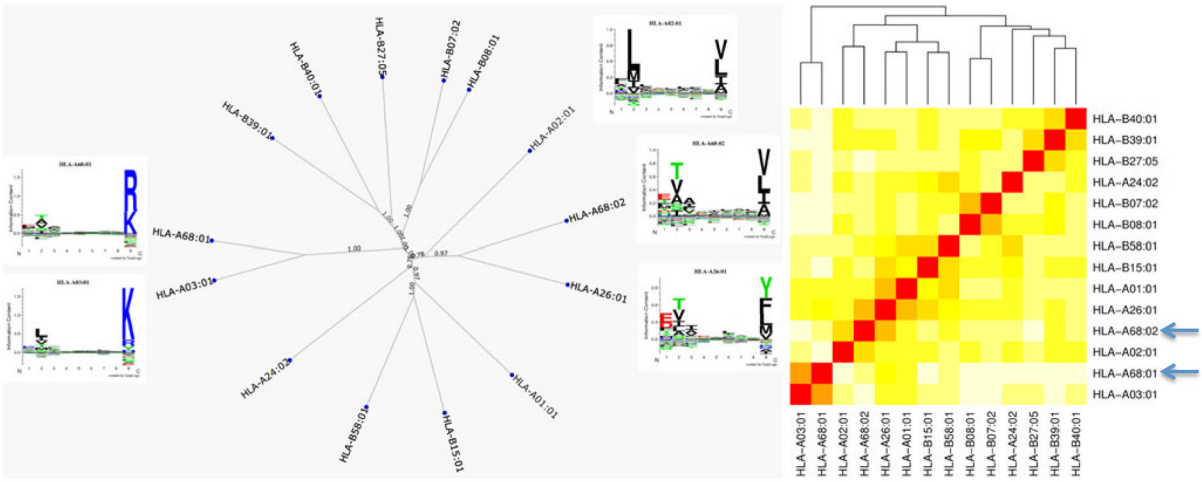

Fig. 6.

Functional clustering of HLA-A*68:01, HLA-A*68:02, and alleles representing the 12 HLA supertypes using *MHCcluster*. The *left panel* shows the tree representation of the clustering and the *right panel* shows the heat map representation. The tree was visualized using the advanced tree-viewer of *MHCcluster*. Logos are included for the HLA-A*68:01, HLA-A*68:02, as well as for the A2, A3, and A24 supertype representatives. HLA-A*68:01 and HLA-A*68:02 are highlighted in the heat map with *arrows*

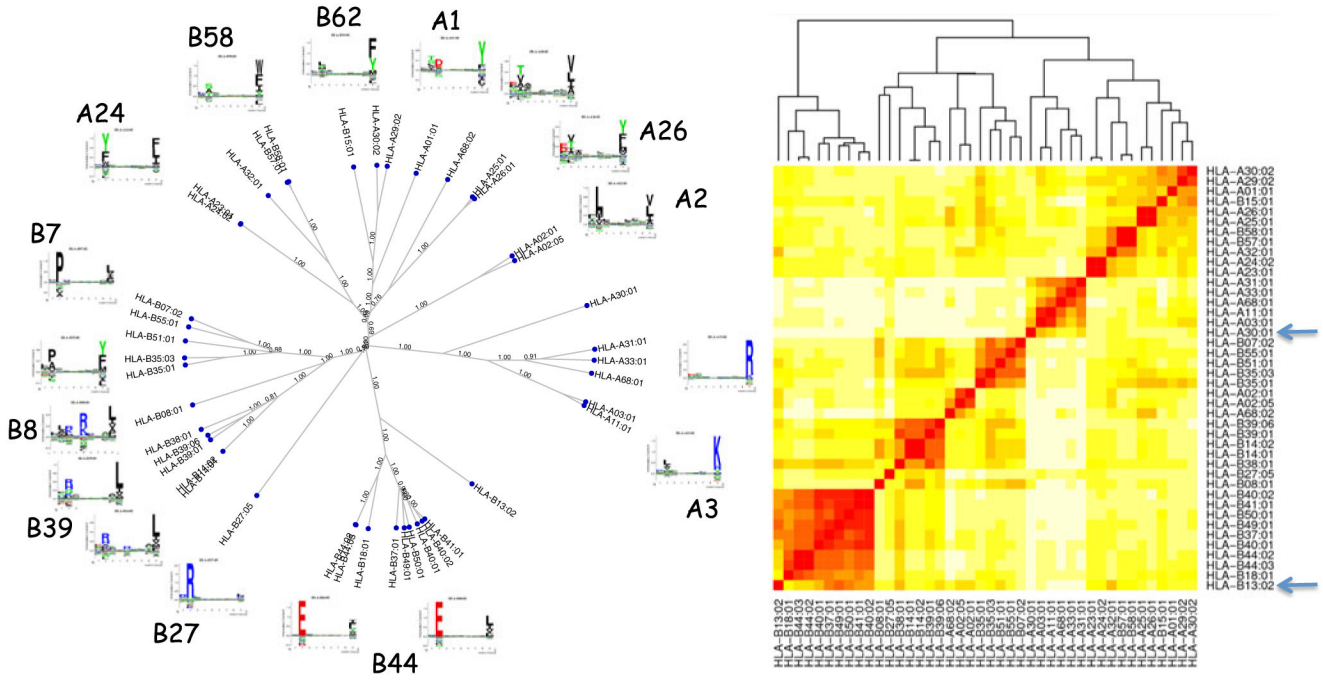

Fig. 7.

Funcional clustering of 42 prevalent HLA-A and B molecules using the *MHCcluster* method. The *left panel* shows the tree representation of the clustering and the *right panel* shows the heat map representation. The tree was visualized using the tree viewer of *MHCcluster*. Logos are included for alleles representing the 12 common supertypes. The location of the two outliers alleles mentioned in the text is highlighted with *arrows* in the heat map

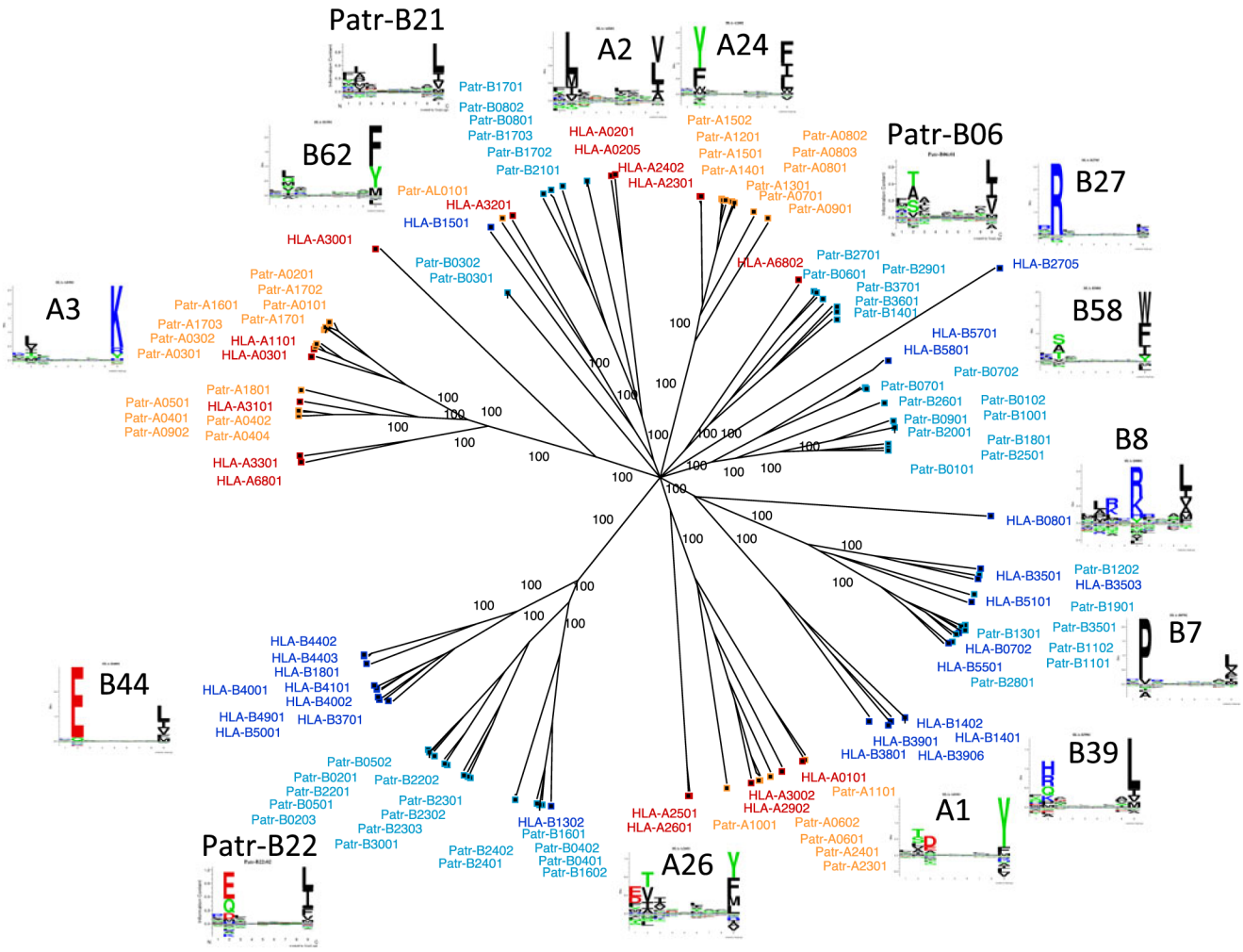

Fig. 8.

Funcional clustering for prevalent HLA, Patr A and B molecules using the *MHCcluster* method. Patr-A alleles are shown in orange, HLA-A alleles in *red*, Patr-B alleles in *light blue*, and HLA-B alleles in *blue*. Logos for the 12 HLA-A and B supertypes are included, as well as logos for the three unique Patr specificities (logos in *shaded gray*). The consensus tree generated by *MHCcluster* was visualized using SplitsTree (Huson and Bryant 2006) and logos were added manually

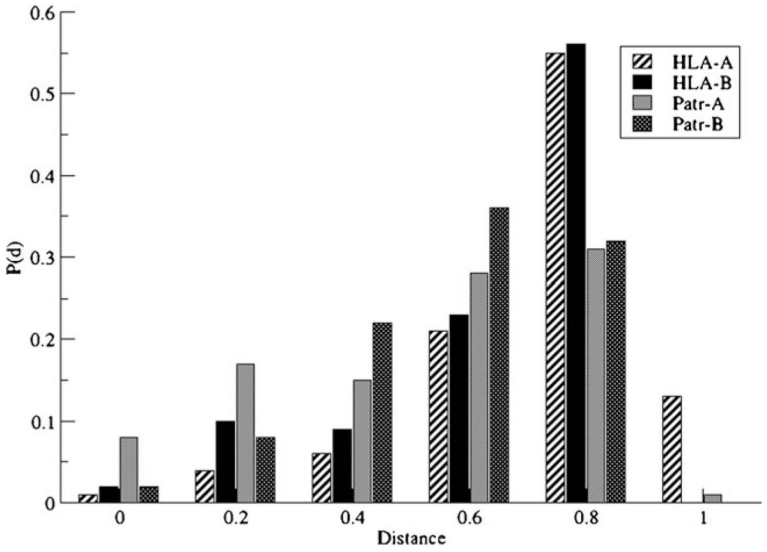

Fig. 9.

Histogram of the pairwise HLA-A HLA-B, Patr-A and Patr-B distances. The histograms were calculated from the 120 MHC molecules included in Fig. 8. P(d) gives the fraction of distances found within a given distance d. Note that the distances are normalized so that the distance between any two MHC molecules falls between 0 and 1