HIV-1 Evolution and Epitope Associations

Sinu Paul
March 18, 2011

Department of Biological Sciences
Kent State University
Kent OH 44242
Outline

• Introduction:
  – HIV-1: Significance, Genome, Evolution
  – Immune system, Epitopes & Epitope vaccine

• Association of CTL epitopes
  – Association rule mining
  – Methods & results

• Association of CTL and T-Helper epitopes
  – Methods & results

• Conclusions
HIV-1 – A shrewd killing machine!!
**AIDS – Acquired Immunodeficiency Syndrome**

- Major cause of death – 2.1 million in 2007 as per UNAIDS
- New infections – 2.5 million in 2007
- Infects CD4+ T-cells (T-Helper cells)
- Destroys immune system

*Image source: http://gsalgeog.blogspot.com/2009_03_01_archive.html*
Why are we so scared?

• Current treatment strategy – **HAART**:
  – Great reduction in mortality but...
  – **Side effects:** Malignancies, Diabetes (Bedimo et al., 2004; Florescu and Kotler, 2007)
  – **Drug resistance, latency** (Ross et al., 2007; Chun et al., 1998)

• **No effective vaccine yet**
  – Modest success of Thai vaccine trial: ALVAC-HIV (vCP1521) prime with an AIDSVAX B/E boost (Rerks-Ngarm, 2009)

• **No cure once infected**
  – One patient cured by stem-cell transplantation (Hutter et al., 2009)
  – **Δ32/Δ32 mutation** in CCR5 receptor
    → **HIV resistant** (Liu et al., 1996)
HIV genome – Tiny but mighty!

- Family: **Retroviridae**

- **Retrovirus** →
  - RNA as genetic material

- 2 strands of positive sense RNA

- RNA to DNA by enzyme **Reverse Transcriptase**

Figure source: http://biology.kenyon.edu/slonec/gene-web/Lentiviral/Lentiviz2.html
Small HIV-1 genome

- 9719 nucleotides
- 9 genes: Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef
- 16 proteins

Figure source: HIV Sequence database by LANL
HIV-1 Origin – A big transition!

HIV:
- Closely related to SIV (Simian Immunodeficiency Virus) – wild chimpanzees, gorilla etc. in West-central Africa (Sharp and Hahn, 2010)
- Multiple cross-species transmissions at different times from different species

SIV:
- Mostly non-pathogenic
- Around 40 different primate species harbor SIV

HIV: HIV-1 & HIV-2
- HIV-2:
  - Concentrated in West Africa
  - 8 groups: A, B, C, D, E, F, G, H
  - A & B transmissible in humans
- HIV-1: Widespread and pathogenic

Photo: Michael L. Wilson. Source: MinnPost.com
SIV & HIV:

Phylogenetic tree made from Pol gene of 57 representative reference sequences showing SIV & HIV
HIV-1 evolution – A killer on the loose!

HIV-1:

- HIV-1 – First identified in 1983 (Barre-Sinoussi et al., 1983)
- 4 groups – M, N, O and recently P group (Plantier et al., 2009)
- Common ancestor of HIV-1 M group dated to 1920s (Korber et al., 2000)
- Several recombinant groups. e.g., BC, BF, HK etc.
- Very rapid evolution → Major challenge in vaccine design

Why HIV is evolving so fast?
HIV-1 – The ultimate evolver

- Error prone Reverse Transcriptase enzyme: 0.2 – 2 mutations per replication cycle (Drake, 1993)

- High replication rate + rapid viral turnover (Perelson et al., 1996):
  - Average total HIV-1 production = 10.3 x 10^9 virions/day
  - Average HIV-1 generation time = 2.6 days

HIV-1 budding from cultured lymphocyte

Image providers: CDC/ C. Goldsmith, P. Feorino, E. L. Palmer, W. R. McManus
**HIV-1: Variability:**

Phylogenetic tree made from Pol gene of 58 representative reference sequences showing HIV-1 variability

- Recombinant sequences

---

**HIV-1: Evolution**

- A
- B
- C
- D
- E
- F
- G
- H
- I
- J
- K
- L
- M

**O group**

**N group**

**M group**

**Recombinant sequences**
HIV-1 global distribution – No one is spared!

Figure source: IAVI Report, August 2003
Immune system & Epitopes -
Playmakers of HIV-1 evolution
Immune system of mammals

- Innate Immunity
  - Relatively simple
  - Non-specific to pathogen
  - Protective barriers
  - Toxins
  - Phagocytic cells

- Adaptive Immunity
  - More complex
  - Antigen/pathogen specific
  - Memory

Figure source: Rabb, H. (2002) Kidney International
Viral infection →

- Class-I MHC molecules recognize viral peptides & interacts with CTLs

Epitopes
Short peptides on the antigens recognized by the immune receptors

Figure source: Sewell et al., 2000
Epitopes

• Critical part of antigen recognition by immune system

• **3 types** based on immune receptors recognizing them:
  – CTL/CD8^+ epitopes
  – T-Helper/CD4^+ epitopes
  – Antibody epitopes

• Some CTL epitopes subject to selective pressure by immune system (Positive selection)

• Some others evolve under purifying selection pressure ➔ Functional and structural constraints
Epitope’s role in HIV-1 evolution

Amino acid changes at epitope regions

Immune system unable to recognize virus

Viral escape

Higher viral fitness

Selection by immune system  

\[ \rightarrow \]

Influence viral evolution

(Price et al., 1997; Goulder and Watkins, 2004; O’Connor et al, 2004)

Figure source: Goulder and Watkins, 2004
Epitope vaccine – A new way to bell the cat?

- Synthetic peptides representing epitopes from pathogens  
  (Jin et al., 2009; Spearman et al., 2009)

- Single or multi-epitope vaccines containing many epitopes  
  (Nehete et al., 2001; Newman et al., 2002)

- Highly conserved epitopes in vaccines:  
  - Target majority of viral variants  
  - Less chance of mutations

- Multiple epitopes:  
  - Less chance of escape  
  - Greater immune response

Figure source: http://goldendaze-ginnie.blogspot.com/2009_04_01_archive.html
Questions

• What epitope combinations are best vaccine candidates?

• Are there (any) highly conserved epitopes (despite high variability of HIV-1)?
  - Wide population coverage
  - Mutations reduce viral fitness

• Do these highly conserved epitopes relate with each other?

• Do they co-occur together in different HIV-1 genomes (subtypes, recombinant forms etc.)?
Association of CTL epitopes
HIV-1 genomic sequence data

- Nucleotide sequences of 9 genes for 62 reference genomes
- From HIV Sequence database by Los Alamos National Laboratory
  - 37 sequences from M group
  - 7 sequences from N & O groups
  - 18 recombinant sequences
- Approx. 4 representative sequences from each class
- Sequences aligned at nucleotide level as per amino acid alignment with ClustalW (Thompson et al., 1995) and manually checked afterwards.
HIV-1 CTL epitopes

• A total of **218 epitopes**: The “best defined HIV CTL epitopes” (Frahm et al., 2007)
• From HIV Immunology database by LANL
• Proven to be immunogenic in humans
• Most evolutionary conserved epitopes selected
• Removed overlapping epitopes with no amino acid difference
• Final set: **189 epitopes**
Identification of associated epitopes:

Association rule mining

- A data mining technique that identifies and describes relationships among items within a data set (Zaki et al., 1997).

- Used in market-basket analysis & also in solving biological questions (Srisawat and Kijsirikul, 2005)
Identification of associated epitopes:

Association rule mining

- Discover novel relationships between CTL epitopes that consistently co-occur together in HIV-1 genomes.

- Apriori algorithm (Agrawal et al., 1993) implemented in program WEKA (Frank et al., 2004)

- Minimum support - 0.75 → Only highly conserved epitopes included

- Confidence - 0.95 → Only very strong associations extracted (Where epitopes co-occur in >95% sequences)
Estimation of relative degree of sequence divergence

• By estimating nucleotide substitution rates

• Pairwise d\text{N} and d\text{S} values (Nei-Gojobori method with Jukes-Cantor correction using \text{MEGA}_4) (Tamura et al., 2007)

• d\text{N} = Nonsynonymous substitutions / Nonsynonymous site

• d\text{S} = Synonymous substitutions / Synonymous site

• Different genomic regions:
  i. Associated epitopes
  ii. Non-associated epitopes
  iii. Non epitope regions
Verification of association rules

• Similar to “Bootstrap” test used in phylogenetics

• 100 “pseudo-sequence sets” containing 62 sequences each created by random selection with replacement from original sequence set.

• This set used as “control-set” to verify presence of association rules found in “reference genomes”
Results

Discovery of CTL epitope associations in HIV-1 reference genomes

- Several highly conserved CTL epitopes found to be associated to each other → They co-occur together in majority of the HIV-1 genomes
Discovery of CTL epitope associations in HIV-1 reference genomes

• A total of 1961 association rules.
• 484 “unique” association rules (rules involving same epitopes collapsed to one).
• Majority of rules involved 2 - 3 epitopes.
• 23 associations involving epitopes from 3 genes: Gag, Pol, Nef.
• These “3-gene associations” are present in >80% of M group genomes.
• Include 22 epitopes from 15 non-overlapping genomic regions.
• Not all highly conserved epitopes take part in association rules.
## Summary of the discovered CTL epitope association rules

<table>
<thead>
<tr>
<th>Number of epitope associations with support &gt;= 0.75 * &amp; confidence &gt;= 0.95</th>
<th>1961</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unique epitope associations*</td>
<td></td>
</tr>
<tr>
<td>Associations with 2 epitopes</td>
<td>46</td>
</tr>
<tr>
<td>Associations with 3 epitopes</td>
<td>217</td>
</tr>
<tr>
<td>Associations with 4 epitopes</td>
<td>153</td>
</tr>
<tr>
<td>Associations with 5 epitopes</td>
<td>59</td>
</tr>
<tr>
<td>Associations with 6 epitopes</td>
<td>9</td>
</tr>
<tr>
<td>Associations with 7 epitopes</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>484</td>
</tr>
<tr>
<td>Unique epitope associations with epitopes from only one gene</td>
<td></td>
</tr>
<tr>
<td>Epitopes from Gag only</td>
<td>9</td>
</tr>
<tr>
<td>Epitopes from Pol only</td>
<td>94</td>
</tr>
<tr>
<td>Epitopes from Nef only</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>103</td>
</tr>
<tr>
<td>Unique epitope associations with epitopes from two genes</td>
<td></td>
</tr>
<tr>
<td>Gag-Pol</td>
<td>329</td>
</tr>
<tr>
<td>Pol-Nef</td>
<td>26</td>
</tr>
<tr>
<td>Nef-Gag</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>358</td>
</tr>
<tr>
<td>Unique epitope associations with epitopes from all three genes (Gag-Pol-Nef)</td>
<td>23</td>
</tr>
</tbody>
</table>
23 CTL epitope association rules that include epitopes from 3 genes
### 22 CTL epitopes involved in the “3-gene” associations

<table>
<thead>
<tr>
<th>Gene</th>
<th>Protein</th>
<th>Non overlapping genomic regions</th>
<th>Amino acid sequence</th>
<th>HLA allele</th>
<th>Amino acid Coordinates</th>
<th>Number of &quot;unique&quot; association rules each epitope is involved</th>
<th>Number of association rules each region is involved</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gag</td>
<td>p24</td>
<td></td>
<td></td>
<td>B*0702</td>
<td>16 24</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B*4001</td>
<td>44 52</td>
<td>127</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B<em>1510, B</em>3901</td>
<td>61 69</td>
<td>214</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B*2705</td>
<td>131 140</td>
<td>7</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B*1501</td>
<td>137 145</td>
<td>195</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cw*18</td>
<td>142 150</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Pol</td>
<td>RT</td>
<td></td>
<td></td>
<td>B*4001</td>
<td>5 12</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A*0301</td>
<td>73 82</td>
<td>188</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A*0301</td>
<td>93 101</td>
<td>124</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B*3301</td>
<td>107 113</td>
<td>38</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B*18</td>
<td>137 146</td>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B*1401</td>
<td>142 149</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B*1501</td>
<td>260 271</td>
<td>112</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A*3002</td>
<td>263 271</td>
<td>114</td>
<td></td>
</tr>
<tr>
<td>RT-RNase</td>
<td></td>
<td></td>
<td></td>
<td>Cw*0802</td>
<td>495 503</td>
<td>149</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B*1503</td>
<td>496 505</td>
<td>153</td>
<td></td>
</tr>
<tr>
<td>RT-Integrase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B*1503</td>
<td>560 8</td>
<td>121</td>
<td>68</td>
</tr>
<tr>
<td>Integrase</td>
<td></td>
<td></td>
<td></td>
<td>B*5701</td>
<td>173 181</td>
<td>52</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A<em>0301, A</em>1101</td>
<td>179 188</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>F*1503</td>
<td>185 194</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cw*0801</td>
<td>260 268</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Nef</td>
<td></td>
<td></td>
<td></td>
<td>B*0801</td>
<td>90 97</td>
<td>52</td>
<td>34</td>
</tr>
</tbody>
</table>
Associated CTL epitope regions more conserved at nucleotide level than other regions

- Comparison of dN and dS values

- Overall, \(d_S > d_N\) in epitope & non-epitope regions \(\rightarrow\) purifying selection

- Associated CTL epitope regions \(\rightarrow\) Lower \(d_N\) & \(d_S\) than other CTL epitopes & non-epitope regions

- Associated CTL epitopes much more conserved at the nucleotide level as well as amino acid level

---

<table>
<thead>
<tr>
<th></th>
<th>(d_N)</th>
<th>(SE)</th>
<th>(d_S)</th>
<th>(SE)</th>
<th>(P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTL epitopes involved in association rules</td>
<td>0.01696</td>
<td>0.00982</td>
<td>0.37794</td>
<td>0.20974</td>
<td>(&lt; 0.01)</td>
</tr>
<tr>
<td>CTL epitopes not involved in association rules</td>
<td>0.12168</td>
<td>0.06814</td>
<td>0.50929</td>
<td>0.18780</td>
<td>(&lt; 0.01)</td>
</tr>
<tr>
<td>Non-epitope regions</td>
<td>0.14698</td>
<td>0.10288</td>
<td>0.53472</td>
<td>0.12572</td>
<td>(&lt; 0.01)</td>
</tr>
</tbody>
</table>
Verification of association rules

• 100 “pseudo-sequence” sets containing a total of 6200 sequences used.

• Essentially *same association rules identified* in the pseudo-sequence sets as that of the reference genome set.

• Result as expected since high support and confidence values already prune away most of the insignificant rules.
Discovery of novel targets for multi-epitope vaccines: Screening of HIV-1 genomes using association rule mining

SINU POUL and HELEN PIONKIVSKA

Department of Biological Sciences, Kent State University, Kent, Ohio 44242, USA

Corresponding author email


The electronic version of this article is the complete one and can be found online at: http://www.retrovirology.com/content/6/1/62

Received: 14 April 2009
Accepted: 6 July 2009
Published: 6 July 2009

© 2009 Poul and Pionkivska; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background

Studies have shown that in the genome of human immunodeficiency virus (HIV-1) regions responsible for interactions with the host's immune system, namely, cytotoxic T-lymphocyte (CTL) epitopes tend to cluster together in relatively conserved regions. On the other hand, “epitope-less” regions or regions with relatively low density of epitopes tend to be more variable. However, very little is known about relationships among epitopes from different genotypes, in other words, whether particular epitopes from different genotypes would occur together in the same viral genome. To identify CTL epitopes in different genotypes that co-occur in HIV genomes, association rule mining was used.

Results

Using a set of 189 best-defined HIV-1 CTL/CD8+ epitopes from 9 different protein-coding genes, as described by Pfaff, Unda & Brandner (2007), we examined the complete genomic sequences of 62 reference HIV sequences (including 12 subtypes and sub-subtypes with approximately 9 representative sequences for each subtype or sub-subtype, and 10 circulating recombinant forms). The results showed that despite inclusion of recombinant sequences that would be expected to break-up associations of epitopes in different genotypes when two different genomes are recombined, there exist particular combinations of epitopes (epitope associations) that occur repeatedly across the worldwide population of HIV-1. For example, Fv epitope LRLGSGDA is found to be significantly associated with epitopes NQQANQL and FLERGKL from Gag and APl, respectively, and this association rule is observed even among circulating recombinant forms.
T-helper epitopes improve vaccine efficiency – increases CTL response (Gram et al., 2009)

In order to evaluate the effect of including a CD4+ T-helper epitope (PADRE) with the CAP01 adjuvants for the induction of a CD8+ T cell response towards a minimal subdominant CD8+ T cell epitope (Vif101) we immunized HLA-A*0201 transgenic HHD mice subcutaneously (s.c.) with Vif101 with or without PADRE. The novel adjuvant CAP01 was used in both immunizations. IFN-γ ELISPOT and a 51Cr-release cytolytic assay was used to evaluate the T cell responses after five days in vitro stimulation.

We found that when immunizing with the CD8 T cell epitope Vif101 together with PADRE, the epitope a high IFN-γ (above 1000 SFU/million cells) towards Vif101 was seen in four out of five animals (Figure 2A) whereas when immunizing with Vif101 alone a

Images from: Gram et al., 2009
Association of CTL and T-Helper epitopes
HIV-1 genomic sequence data

- 90 HIV-1 reference genomes containing 9 protein-coding genes
- From HIV Sequence database by LANL
- 40 sequences from M group
- 7 sequences from N & O groups
- 43 recombinant sequences
- Approx. 4 representative sequences from each class
- Nucleotide and amino acid sequence alignments were downloaded.
HIV-1 Epitopes

- A total of 606 epitopes: 229 CTL, 296 T-Helper and 81 Antibody epitopes
- From HIV Immunology database by LANL
- Proven to be immunogenic in humans
- Epitopes present in > 75% of the reference sequences selected for association rule mining.
- Removed overlapping epitopes with no amino acid difference
- Final set: 44 epitopes (32 CTL, 10 Th and 2 Ab) from 4 genes (Gag, Pol, Env and Nef)
### Overview of the number of epitopes used in the analysis

<table>
<thead>
<tr>
<th>Gene</th>
<th>Protein</th>
<th>Total no. of epitopes</th>
<th>Highly conserved epitopes</th>
<th>No. of associated epitopes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>CTL</td>
<td>Th</td>
<td>Ab</td>
</tr>
<tr>
<td>Gag</td>
<td>p17</td>
<td>18</td>
<td>32</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>p24</td>
<td>42</td>
<td>88</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>p2p7p1p6</td>
<td>6</td>
<td>18</td>
<td>24</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>66</td>
<td>138</td>
<td>1</td>
</tr>
<tr>
<td>Pol</td>
<td>Gag-Pol</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Protease</td>
<td>8</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RT</td>
<td>39</td>
<td>20</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>RT-Integrase</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Integrase</td>
<td>12</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>61</td>
<td>32</td>
<td>3</td>
</tr>
<tr>
<td>Vif</td>
<td></td>
<td>9</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Vpr</td>
<td></td>
<td>7</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Tat</td>
<td></td>
<td>4</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Rev</td>
<td></td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Vpu</td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Env</td>
<td></td>
<td>40</td>
<td>82</td>
<td>75</td>
</tr>
<tr>
<td>Nef</td>
<td></td>
<td>37</td>
<td>24</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>229</td>
<td>296</td>
<td>81</td>
</tr>
</tbody>
</table>
Identification of associated epitopes

- Association rule mining
- Apriori algorithm
- Minimum support - 0.75 and confidence - 0.95

Estimation of sequence divergence

- Pairwise dN and dS values (Nei-Gojobori method with Jukes-Cantor correction using MEGA4)
- For different genomic regions:
  i. Associated epitopes
  ii. Non-associated epitopes
  iii. Variable epitopes
  iv. Non epitope regions
Verification of association rules

• A set of 978 HIV-1 genomes
• From 2008 web alignment of HIV Sequence Database
• 650 HIV-1 group M sequences
• 22 N & O sequences
• 306 recombinant sequences
• Analyzed the presence of the association rules discovered in reference genomes
Results

Discovery of CTL-T-Helper epitope associations in HIV-1 reference genomes

- 6142 unique association rules
- The rules involved 2 – 9 epitopes with majority having 3 – 5 epitopes

Distribution of unique association rules according to genes and no. of epitopes involved in each rule

<table>
<thead>
<tr>
<th>Association rules with 2 epitopes</th>
<th>Gag only</th>
<th>Pol only</th>
<th>Nef only</th>
<th>Gag Pol</th>
<th>Gag Env</th>
<th>Gag Nef</th>
<th>Pol Env</th>
<th>Pol Nef</th>
<th>Gag Pol Nef</th>
<th>Total*</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 epitopes</td>
<td>104</td>
<td>160</td>
<td>0</td>
<td>768</td>
<td>1</td>
<td>33</td>
<td>0</td>
<td>23</td>
<td>56</td>
<td>1145</td>
</tr>
<tr>
<td>4 epitopes</td>
<td>108</td>
<td>135</td>
<td>0</td>
<td>1699</td>
<td>0</td>
<td>29</td>
<td>0</td>
<td>23</td>
<td>104</td>
<td>2098</td>
</tr>
<tr>
<td>5 epitopes</td>
<td>73</td>
<td>47</td>
<td>0</td>
<td>1551</td>
<td>0</td>
<td>11</td>
<td>0</td>
<td>4</td>
<td>33</td>
<td>1719</td>
</tr>
<tr>
<td>6 epitopes</td>
<td>29</td>
<td>6</td>
<td>0</td>
<td>753</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>793</td>
</tr>
<tr>
<td>7 epitopes</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>211</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>216</td>
</tr>
<tr>
<td>8 epitopes</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>31</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>31</td>
</tr>
<tr>
<td>9 epitopes</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Total*</td>
<td>365</td>
<td>372</td>
<td>1</td>
<td>5070</td>
<td>4</td>
<td>80</td>
<td>1</td>
<td>53</td>
<td>196</td>
<td>6142</td>
</tr>
</tbody>
</table>
• 137 “Multi-type, multi-gene” association rules – 2 types of epitopes (CTL & Th) from 3 genes (Gag, Pol and Nef)
Example of a “multi-type, multi-gene” association rule
### 21 epitopes involved in “2T-3G” rules

<table>
<thead>
<tr>
<th>Gene</th>
<th>Protein</th>
<th>Non-overlapping genomic regions</th>
<th>Epitope sequence</th>
<th>amino acid coordinates</th>
<th>Type of epitope</th>
<th>Epitopes involved in 2T-3G epitopes</th>
<th>Non-overlapping genomic regions of 2T-3G epitopes</th>
<th>Number of “unique” associations each epitope is involved in</th>
<th>HLA allele/Mab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gag</td>
<td>p17</td>
<td></td>
<td>WASHKRES</td>
<td>36 44</td>
<td>CTL</td>
<td>0</td>
<td></td>
<td></td>
<td>B<em>5101, B</em>5801, B*53</td>
</tr>
<tr>
<td></td>
<td>p24</td>
<td></td>
<td>HPRILHAYV</td>
<td>16 24</td>
<td>CTL</td>
<td>1</td>
<td></td>
<td></td>
<td>B<em>0702, B</em>41</td>
</tr>
<tr>
<td></td>
<td>p24</td>
<td></td>
<td>FSPVUPVAF</td>
<td>52 40</td>
<td>CTL</td>
<td>8</td>
<td></td>
<td></td>
<td>B*57</td>
</tr>
<tr>
<td></td>
<td>p24</td>
<td></td>
<td>EVPSAASL</td>
<td>55 43</td>
<td>CTL</td>
<td>1</td>
<td></td>
<td></td>
<td>A<em>5801, A</em>6801, B<em>1501, B</em>0001</td>
</tr>
<tr>
<td></td>
<td>p24</td>
<td></td>
<td>SGAVTPQGL</td>
<td>44 27</td>
<td>CTL</td>
<td>1</td>
<td></td>
<td></td>
<td>A<em>3001, B</em>4601</td>
</tr>
<tr>
<td></td>
<td>p24</td>
<td></td>
<td>GHSLAACKG</td>
<td>61 69</td>
<td>CTL</td>
<td>2</td>
<td></td>
<td></td>
<td>A<em>0201, B</em>1510, B<em>08, B</em>3501</td>
</tr>
<tr>
<td></td>
<td>p24</td>
<td></td>
<td>EPQVISDDAGT</td>
<td>98 107</td>
<td>TH</td>
<td>3</td>
<td></td>
<td></td>
<td>DQ7</td>
</tr>
<tr>
<td></td>
<td>p24</td>
<td></td>
<td>IYKRWILGLNUNQV</td>
<td>129 143</td>
<td>TH</td>
<td>1</td>
<td></td>
<td></td>
<td>1167</td>
</tr>
<tr>
<td></td>
<td>p24</td>
<td></td>
<td>KRWLGLNK</td>
<td>131 140</td>
<td>CTL</td>
<td>1</td>
<td></td>
<td></td>
<td>B<em>2703, B</em>2705, B<em>35, B</em>0101</td>
</tr>
<tr>
<td></td>
<td>p24</td>
<td></td>
<td>KRWLGLNKLK</td>
<td>131 140</td>
<td>CTL</td>
<td>1</td>
<td></td>
<td></td>
<td>DR1, DRB1<em>0101, DRB1</em>0301, DRB1<em>0403, DRB1</em>0701, DRB1<em>0802, DRB1</em>0901, DRB1<em>1101, DRB1</em>1201, DRB1<em>1302, DRB1</em>1401, DRB1<em>0101, DRB1</em>0103</td>
</tr>
<tr>
<td></td>
<td>p24</td>
<td></td>
<td>WILGLKINGMYMP</td>
<td>133 147</td>
<td>TH</td>
<td>3</td>
<td></td>
<td></td>
<td>B*3501</td>
</tr>
<tr>
<td></td>
<td>p24</td>
<td></td>
<td>GLNIMVNY</td>
<td>157 143</td>
<td>TH</td>
<td>3</td>
<td></td>
<td></td>
<td>B*0801</td>
</tr>
<tr>
<td></td>
<td>p24</td>
<td></td>
<td>GLNIMVNY</td>
<td>157 143</td>
<td>TH</td>
<td>3</td>
<td></td>
<td></td>
<td>B*3501</td>
</tr>
<tr>
<td></td>
<td>p24</td>
<td></td>
<td>LINKIMVNYTPSRLD</td>
<td>138 152</td>
<td>TH</td>
<td>15</td>
<td></td>
<td></td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>p24</td>
<td></td>
<td>VRGMYPVPH</td>
<td>142 150</td>
<td>CTL</td>
<td>46</td>
<td></td>
<td></td>
<td>C*18</td>
</tr>
<tr>
<td></td>
<td>p24</td>
<td></td>
<td>KEPKPRETV</td>
<td>157 165</td>
<td>TH</td>
<td>4</td>
<td></td>
<td></td>
<td>DQ5</td>
</tr>
<tr>
<td></td>
<td>p24</td>
<td></td>
<td>CPFQKEKCC</td>
<td>42 50</td>
<td>CTL</td>
<td>9</td>
<td></td>
<td></td>
<td>B*14</td>
</tr>
<tr>
<td></td>
<td>p24</td>
<td></td>
<td>TERQANPL</td>
<td>64 71</td>
<td>CTL</td>
<td>26</td>
<td></td>
<td></td>
<td>B<em>1801, B</em>4001, B<em>4003, B</em>4402, B*4403</td>
</tr>
<tr>
<td>Pol</td>
<td>p9</td>
<td></td>
<td>INQTPTPVHL</td>
<td>76 84</td>
<td>CTL</td>
<td>1</td>
<td></td>
<td></td>
<td>A<em>0201, A</em>0202, A<em>0203, A</em>6802</td>
</tr>
<tr>
<td></td>
<td>p9</td>
<td></td>
<td>RT</td>
<td>5 12</td>
<td>CT</td>
<td>17</td>
<td></td>
<td></td>
<td>B*4001</td>
</tr>
<tr>
<td></td>
<td>p9</td>
<td></td>
<td>GRLVQKWPL</td>
<td>18 26</td>
<td>CTL</td>
<td>6</td>
<td></td>
<td></td>
<td>B*0801, B3</td>
</tr>
<tr>
<td></td>
<td>p9</td>
<td></td>
<td>KLVDPNFLK</td>
<td>73 82</td>
<td>CTL</td>
<td>5</td>
<td></td>
<td></td>
<td>1354 *0301</td>
</tr>
<tr>
<td></td>
<td>p9</td>
<td></td>
<td>GNPQAGLLK</td>
<td>93 101</td>
<td>CTL</td>
<td>6</td>
<td></td>
<td></td>
<td>A<em>0301, A</em>11</td>
</tr>
<tr>
<td></td>
<td>p9</td>
<td></td>
<td>FVQYGQAY</td>
<td>107 113</td>
<td>CTL</td>
<td>7</td>
<td></td>
<td></td>
<td>971</td>
</tr>
<tr>
<td></td>
<td>p9</td>
<td></td>
<td>NETPQEYQY</td>
<td>137 146</td>
<td>CTL</td>
<td>5</td>
<td></td>
<td></td>
<td>B<em>1801, B</em>4001, B<em>4003, B</em>4402, B*4403</td>
</tr>
<tr>
<td></td>
<td>p9</td>
<td></td>
<td>KEQFYNL</td>
<td>142 159</td>
<td>CTL</td>
<td>31</td>
<td></td>
<td></td>
<td>B*1401</td>
</tr>
<tr>
<td></td>
<td>p9</td>
<td></td>
<td>LVQKILAWAQY</td>
<td>260 271</td>
<td>CTL</td>
<td>8</td>
<td></td>
<td></td>
<td>B*1501</td>
</tr>
<tr>
<td></td>
<td>p9</td>
<td></td>
<td>KLWMAIAGY</td>
<td>263 271</td>
<td>CTL</td>
<td>1117</td>
<td></td>
<td></td>
<td>A*0002</td>
</tr>
<tr>
<td></td>
<td>p9</td>
<td></td>
<td>WEFVSHPPLTKVLY QY</td>
<td>414 428</td>
<td>TH</td>
<td>85</td>
<td></td>
<td></td>
<td>1376</td>
</tr>
<tr>
<td>Integrase</td>
<td></td>
<td></td>
<td>GAETFYDGA</td>
<td>436 445</td>
<td>CTL</td>
<td>11</td>
<td></td>
<td></td>
<td>A*6802</td>
</tr>
<tr>
<td></td>
<td>Integrase</td>
<td></td>
<td>JGVTSAAYL</td>
<td>405 503</td>
<td>CTL</td>
<td>9</td>
<td></td>
<td></td>
<td>471</td>
</tr>
<tr>
<td></td>
<td>Integrase</td>
<td></td>
<td>GGRQVAAYL</td>
<td>496 504</td>
<td>CTL</td>
<td>7</td>
<td></td>
<td></td>
<td>837</td>
</tr>
<tr>
<td></td>
<td>Integrase</td>
<td></td>
<td>KT</td>
<td>509 509</td>
<td>CTL</td>
<td>10</td>
<td></td>
<td></td>
<td>B*5703</td>
</tr>
<tr>
<td></td>
<td>Integrase</td>
<td></td>
<td>LKTAQMASTRNFKK</td>
<td>172 186</td>
<td>CTL</td>
<td>11</td>
<td></td>
<td></td>
<td>1172</td>
</tr>
<tr>
<td></td>
<td>Integrase</td>
<td></td>
<td>KTAAQMAV</td>
<td>173 181</td>
<td>CTL</td>
<td>113</td>
<td></td>
<td></td>
<td>1129</td>
</tr>
<tr>
<td></td>
<td>Integrase</td>
<td></td>
<td>KTAAQMASTRNFKK</td>
<td>173 187</td>
<td>TH</td>
<td>113</td>
<td></td>
<td></td>
<td>1172</td>
</tr>
<tr>
<td></td>
<td>Integrase</td>
<td></td>
<td>AVFPHFNFRRK</td>
<td>179 188</td>
<td>CTL</td>
<td>5</td>
<td></td>
<td></td>
<td>631 <em>0301, A</em>1101</td>
</tr>
<tr>
<td></td>
<td>Integrase</td>
<td></td>
<td>PKEFSGGIGGY</td>
<td>185 194</td>
<td>CTL</td>
<td>3</td>
<td></td>
<td></td>
<td>195</td>
</tr>
<tr>
<td></td>
<td>Integrase</td>
<td></td>
<td>PFRKAKK</td>
<td>260 282</td>
<td>CTL</td>
<td>12</td>
<td></td>
<td></td>
<td>B*1503</td>
</tr>
<tr>
<td></td>
<td>Integrase</td>
<td></td>
<td>RFAKQDPY</td>
<td>263 271</td>
<td>CTL</td>
<td>6</td>
<td></td>
<td></td>
<td>B*1503</td>
</tr>
<tr>
<td>Env</td>
<td>p26</td>
<td></td>
<td>9PPHHCAPA</td>
<td>212 221</td>
<td>Ab</td>
<td>0</td>
<td></td>
<td></td>
<td>1103</td>
</tr>
<tr>
<td></td>
<td>p26</td>
<td></td>
<td>IKGI</td>
<td>420 423</td>
<td>Ab</td>
<td>5</td>
<td></td>
<td></td>
<td>B*51</td>
</tr>
<tr>
<td></td>
<td>p26</td>
<td></td>
<td>VGEVRQPD</td>
<td>66 73</td>
<td>TH</td>
<td>13</td>
<td></td>
<td></td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>p26</td>
<td></td>
<td>RPQRPPEPM</td>
<td>71 79</td>
<td>CTL</td>
<td>13</td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>p26</td>
<td></td>
<td>FLEKKSQG</td>
<td>90 97</td>
<td>CTL</td>
<td>14</td>
<td></td>
<td></td>
<td>258</td>
</tr>
<tr>
<td>Nef</td>
<td>p26</td>
<td></td>
<td>9PPHHCAPA*</td>
<td>212 221</td>
<td>Ab</td>
<td>0</td>
<td></td>
<td></td>
<td>1103</td>
</tr>
<tr>
<td></td>
<td>p26</td>
<td></td>
<td>IKGI</td>
<td>420 423</td>
<td>Ab</td>
<td>5</td>
<td></td>
<td></td>
<td>B*51</td>
</tr>
<tr>
<td></td>
<td>p26</td>
<td></td>
<td>VGEVRQPD</td>
<td>66 73</td>
<td>TH</td>
<td>13</td>
<td></td>
<td></td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>p26</td>
<td></td>
<td>RPQRPPEPM</td>
<td>71 79</td>
<td>CTL</td>
<td>13</td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>p26</td>
<td></td>
<td>FLEKKSQG</td>
<td>90 97</td>
<td>CTL</td>
<td>14</td>
<td></td>
<td></td>
<td>258</td>
</tr>
</tbody>
</table>
Epitope-associations in the reference genome are representative of the global HIV-1 population

- Most of the epitope-associations present in majority of global HIV-1 genome set

- out of 137 2T-3G epitope associations, 134 association rules present in >70% of the HIV-1 genomes (>685 sequences)

- 20 of the 21 epitopes involved in “2T-3G” associations present in >85% sequences of the global set

- Indicates the importance of these highly conserved & associated epitopes.
Associated epitope regions highly conserved at nucleotide level as well

- Overall, \( dS \gg dN \) – indicates purifying selection plays a significant role in the evolution of HIV-1 including evolution of the epitope regions

- Associated epitopes have significantly smaller \( dN \) and \( dS \) values than those at other categories of sites, including non-epitopes

- High degree of sequence conservation exist not only at the amino acid level, but also at the nucleotide level in these associated regions

<table>
<thead>
<tr>
<th></th>
<th>( dN )</th>
<th>( SE ) #</th>
<th>( dS )</th>
<th>( SE )</th>
<th>( P)-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Associated epitopes</td>
<td>0.01062</td>
<td>0.00952</td>
<td>0.20969</td>
<td>0.07091</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Non-associated epitopes</td>
<td>0.02387</td>
<td>0.02537</td>
<td>0.24220</td>
<td>0.12666</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Not included epitopes</td>
<td>0.10532</td>
<td>0.01277</td>
<td>0.29085</td>
<td>0.04305</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Non-epitopes</td>
<td>0.09793</td>
<td>0.01653</td>
<td>0.27329</td>
<td>0.04665</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>
Frequent associations between CTL and T-Helper epitopes in HIV-1 genomes and implications for multi-epitope vaccine designs

Sinu Paul and Helen Piontkivska
Department of Biological Sciences, Kent State University, Kent, Ohio, 44242, USA

Abstract

Background

Epitope vaccines have been suggested as a strategy to counteract viral escape and development of drug resistance. Multiple studies have shown that Cytotoxic T-Lymphocyte (CTL) and T-Helper (Th) epitopes can generate strong immune responses in Human Immunodeficiency Virus (HIV-1). However, not much is known about the relationship among different types of HIV epitopes, particularly those epitopes that can be considered potential candidates for inclusion in the multi-epitope vaccines.

Results

In this study we used association rule mining to examine relationship between different types of epitopes (CTL, Th and antibody epitopes) from nine protein-encoding HIV-1 genes to identify strong associations as potent multi-epitope vaccine candidates. Our results revealed 137 association rules that were consistently present in the majority of reference and non-reference HIV-1 genomes and included epitopes of two different types (CTL and Th) from three different genes (Gag, Pol and Rev). These rules involved 14 non-overlapping epitope regions that frequently co-occurred despite high mutation and recombination rates, including in genomes of circulating recombinant forms. These epitope regions were also highly conserved at both the amino acid and nucleotide levels indicating strong purifying selection driven by functional and/or structural constraints and hence, the diminished likelihood of successful escape mutations.
• A package of several programs for the analysis of linear epitope sequences
In a nutshell

- Identified several highly conserved CTL and T-Helper epitopes that co-occur in the majority of HIV-1 genomes found around the world.

- These highly conserved & associated epitopes can be considered as potent candidates for a multi-epitope vaccine against HIV-1.

- Mutations in one epitope will not completely diminish the vaccine efficacy.

- Escape due to mutations in these highly conserved epitopes is unlikely to happen due to structural/functional constraints.

- Epitopes of different types can induce broader immune response.
What next?

- Pattern of amino acid changes in associated epitopes → Molecular coevolution of epitope regions.
- Nature of epitope evolution in acute and chronic infections.
- Delineate selective pressures elicited by different branches of immune system like Cell mediated and Humoral and their evolutionary implications.
Acknowledgements

• Dr. Helen Piontkivska

• Labmates:
  – Satish Perikala
  – Reeba Paul
  – Abdul Khaleg

• Friends & Family
Thank you