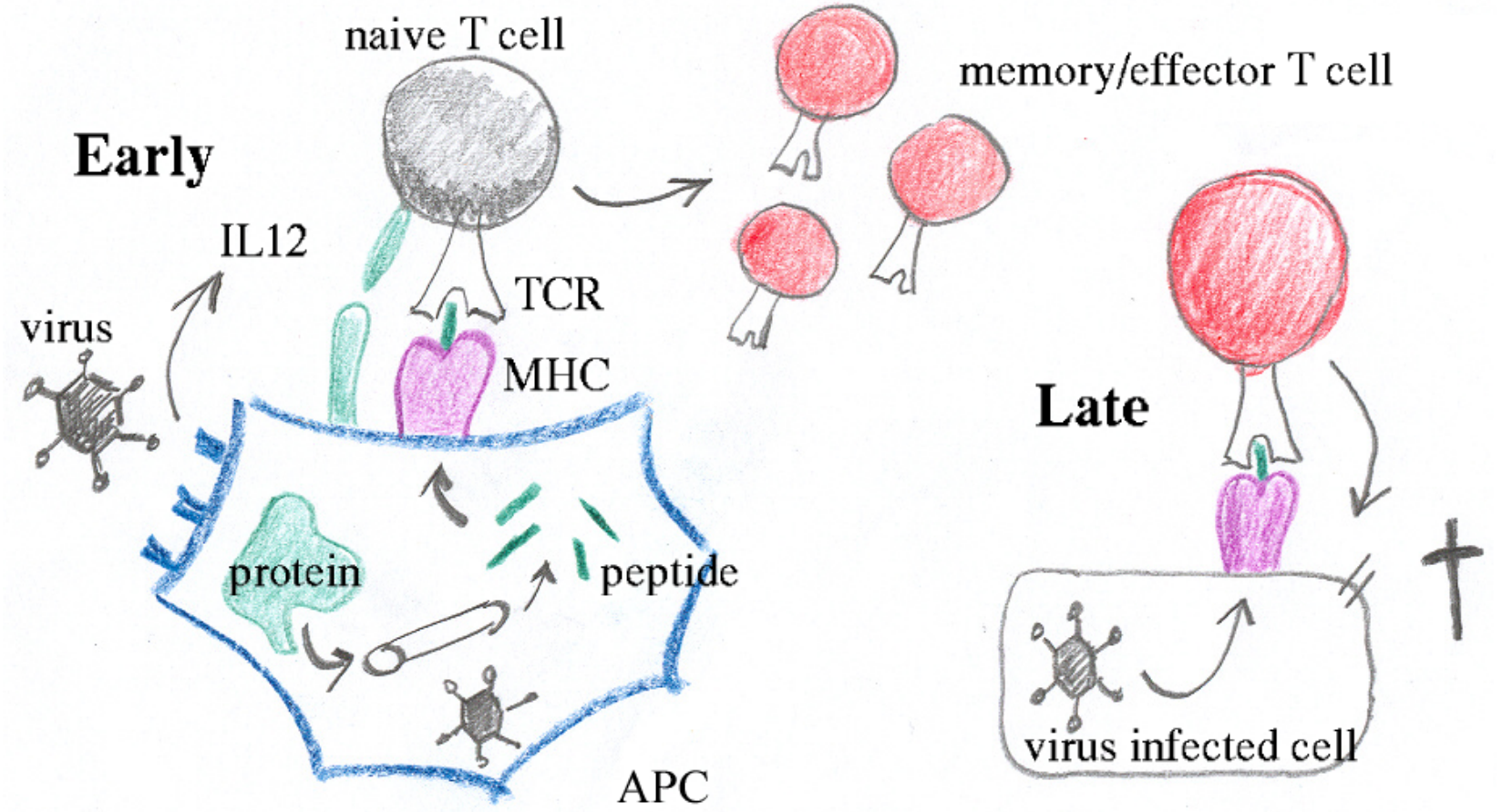
# MHC Polymorphism and Peptide Diversity

Rob J. de Boer
Theoretical Biology, UU

# Sample a few peptides (9-mers) from a pathogen




Because MHC is polymorphic we all draw different samples

## Foreign peptides need to be detected

- peptides from pathogen different from self peptides
- need T cells specific for peptides from pathogen
- peptides from pathogen compete with self peptides
- pathogens evolve escape mutants that fail to be presented by MHC or be detected by T cells

## Diversity and Polymorphism in the immune system

**Diversity of lymphocytes**

- Huge diversity **within a host**
- At least $10^8$ different T & B cell clones

**Diversity of self peptides**

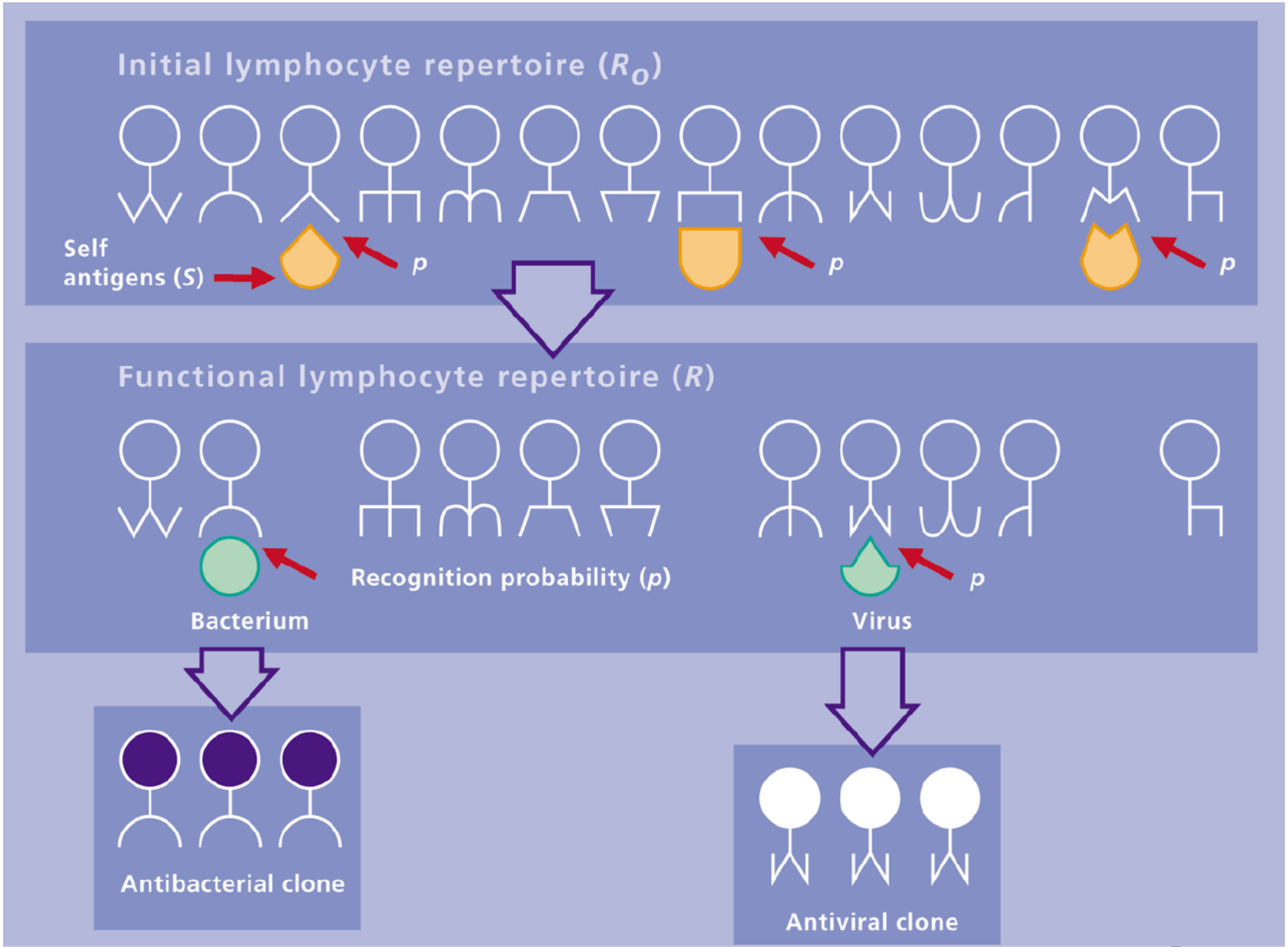
- $3 \times 10^4$ self proteins is $10^7$ unique 9-mers

**Polymorphism of MHC**

- Within a host limited number of loci (genes)

$\rightarrow$ only 6 class I and 12 class II molecules

- Within a **population** $>$ 100 alleles per locus


Initial lymphocyte repertoire ($R_O$)
Self antigens ($S$)
p
p
p
Functional lymphocyte repertoire ($R$)
Recognition probability ($p$)
Bacterium
p
Virus
Antibacterial clone
Antiviral clone

## Diversity of lymphocytes and self peptides

- **De Boer & Perelson, PRSL B 1993**:
  Lymphocytes specific to prevent deletion by self tolerance:

$$P_i = 1 - (1-p)^{R_0(1-p)^S} \quad \rightarrow \quad p \simeq \frac{1}{S}$$

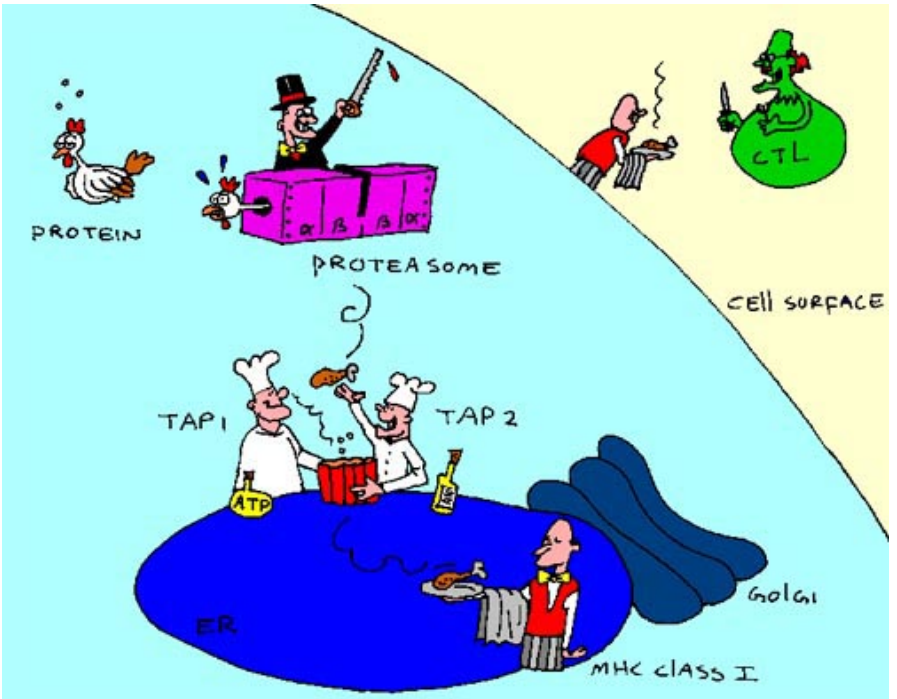
  Specificity seems determined by number of self epitopes
  Large repertoire required if T cells are specific
- **Borghans et al. J. Immunol. 1999**:
  Lymphocytes instructed for certain type of response should not crossreact: be as specific as possible
- **Borghans & De Boer, Int. Immunol. 2002**:
  Memory harmful when lymphocytes too crossreactive

## Conclusion: lymphocyte diversity

Lymphocyte specificity, and hence diversity, for a large part determined by the diversity of self peptides.

To reliably store the innate instruction/decision into effector cells, and to recall that memory during subsequent decisions for related pathogens, lymphocytes have to be sufficiently specific.

# Short (9-mer) peptides exposed on class I MHC




What is the diversity of self peptides?

How much information is present in a 9-mer?

From the webpage of J. Neefjes (NKI)

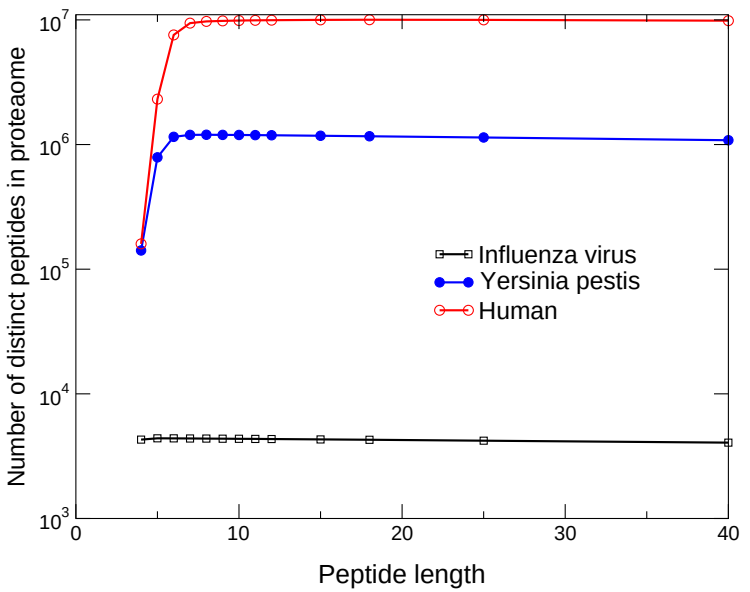
## Number and Uniqueness of self peptides on class I MHC

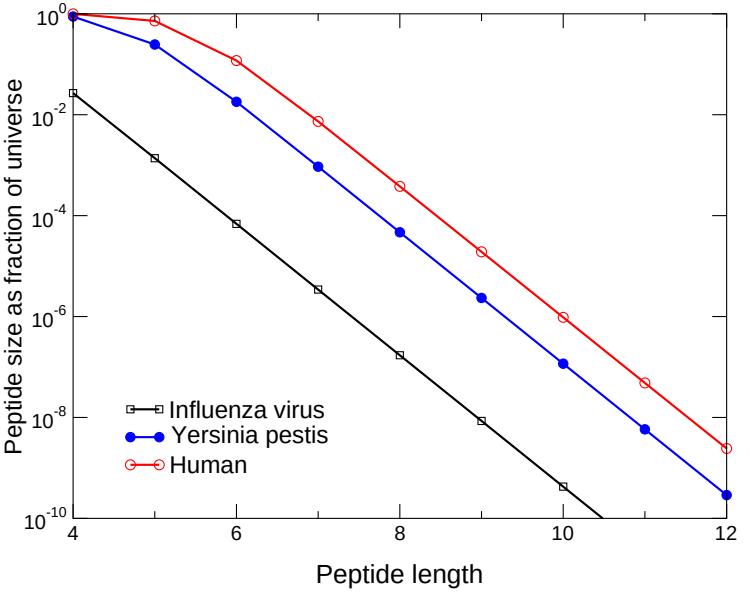
Count the number of unique peptides of length $6, 7, \ldots$ in human and pathogen proteomes.

- human self is $3 \times 10^4$ proteins of $10^7$ distinct 9-mers
- 75% occurs only once: most 9-mers are unique

Burroughs, De Boer & Keşmir, Immunogenetics, 2004

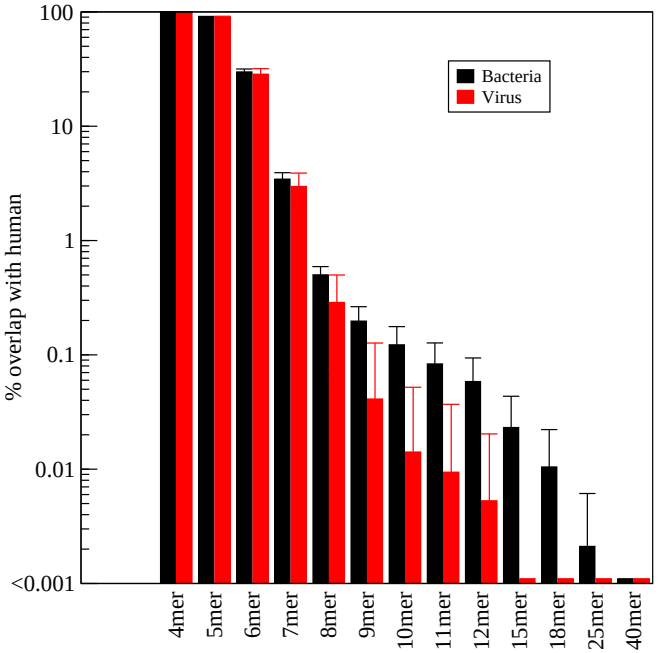
# Size of self



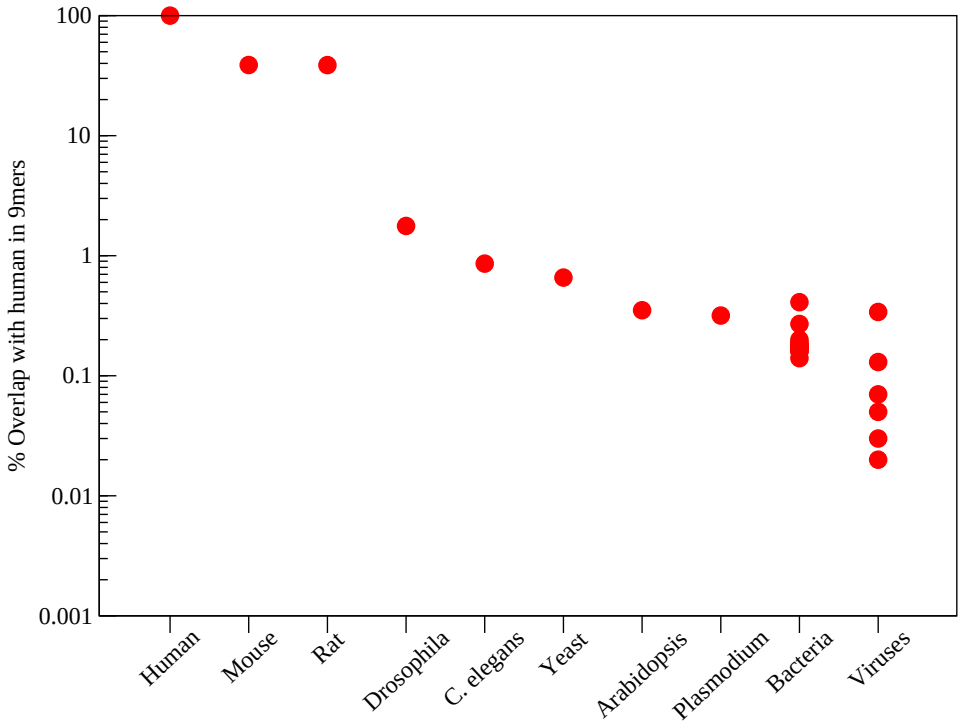

$10^7 \ll 20^9$: self is a small fraction of peptide space: overlaps with other "selfs" expected to be small.

# Overlaps human and 14 bacteria and 17 viruses



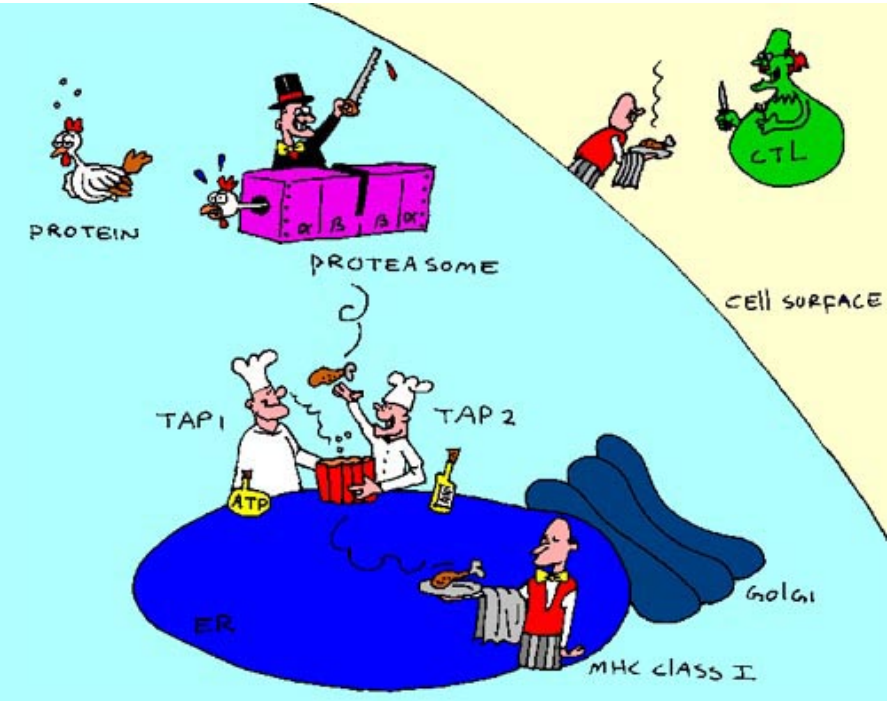

9-mers provide enough information to discriminate:
average overlap human and pathogens <1%
(Overlaps between unrelated bacteria also 1%)

# Overlaps with human depend on evolutionary distance




Overlaps with mouse and rat similar to repeats in human

# TAP and proteasome filter the set of presented peptides




Do they discrimate self from non-self?

## Processing and presentation of 9-mers

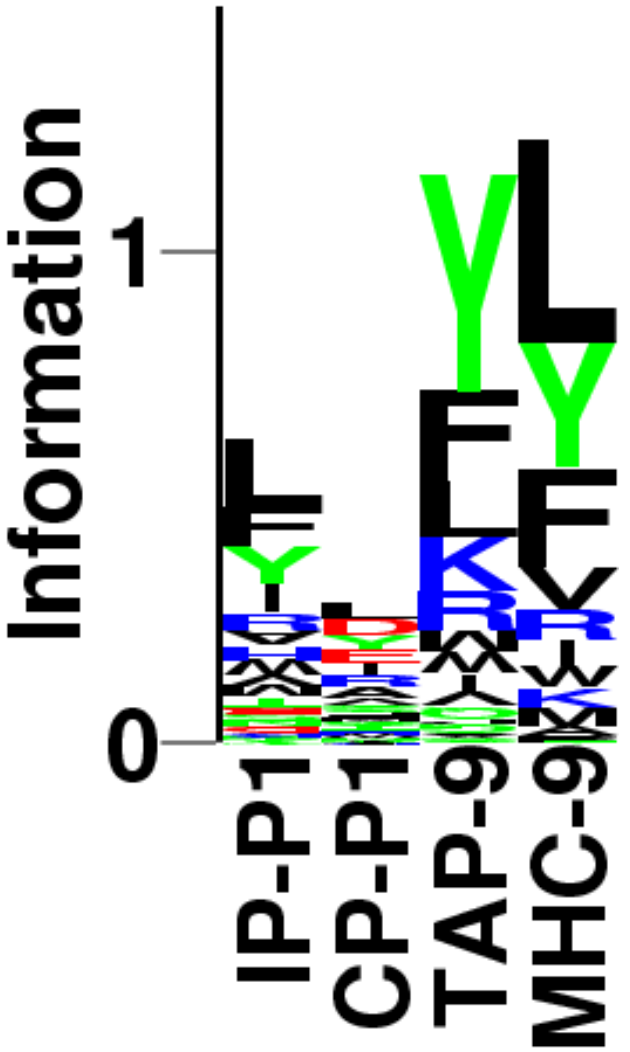
| | Proteasome | | TAP | | MHC$^{A*0204}$ | |
|---|---|---|---|---|---|---|
| | **S %** | **NS %** | **S %** | **NS %** | **S %** | **NS %** |
| All 9-mers | **33.9** | **33.6** | **58.5** | **61.7** | **2.6** | **3.6** |
| Cleaved | | | **71.6** | **75.7** | **5.5** | **7.0** |
| Translocated | | | | | **3.8** | **4.5** |
| Processed | | | | | **6.7** | **8.1** |

MHC and TAP have preference for non-self peptides (due to different aa usage between human and foreign)

TAP has a preference for cleaved peptides

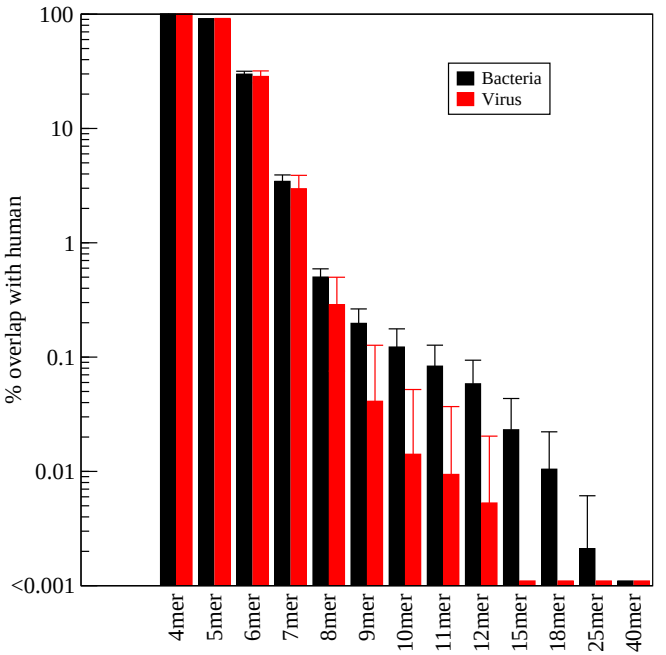
MHC has a preference for processed peptides

# Co-evolution of specificities at the C terminal



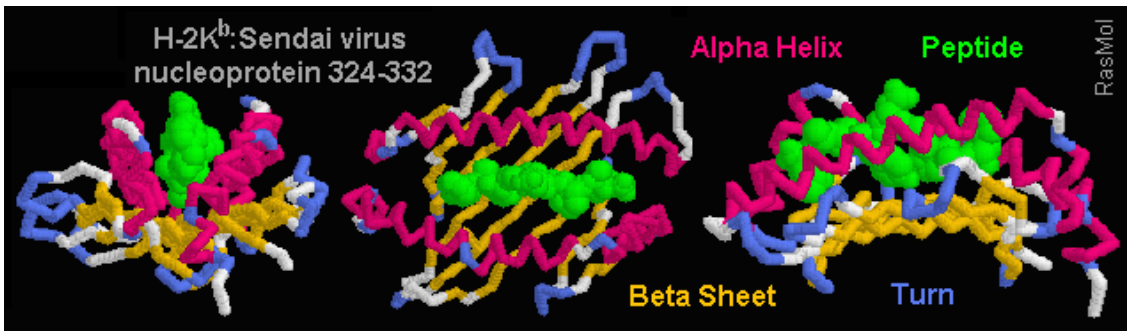

Leucine, tyrosine, fenylanaline, and arginine are preferred at C terminal

# Back to the overlaps




9-mer: $< 1$ %, 7-mer: 3 %

# Anchor residues are not visible to T cell




Not all information from 9-mer is available to discriminate between peptides

From: `http://www.umass.edu/microbio/rasmol`

## Overlaps when HLA anchor residues are not visible

| | 9-mer overlap | HLA-A*0201 overlap | HLA-A*0201 presented | HLA-A*0204 overlap | HLA-A*0204 presented |
|---|---|---|---|---|---|
| human | — | — | 3.7% | — | 6.7% |
| bact. | $0.2 \pm 0.1$ | $0.3 \pm 0.1$ | $4.2 \pm 0.2$ | $0.4 \pm 0.1$ | $7.8 \pm 0.9$ |
| virus | $0.04 \pm 0.1$ | $0.3 \pm 1.0$ | $4.0 \pm 1.0$ | $0.2 \pm 0.5$ | $7.4 \pm 1.0$ |

Predict processed & presented peptides for HLA alleles

Remove two anchor residues: 9-mer $\rightarrow$ 7-mer,

and count overlap on the 7-mer level of the T cells

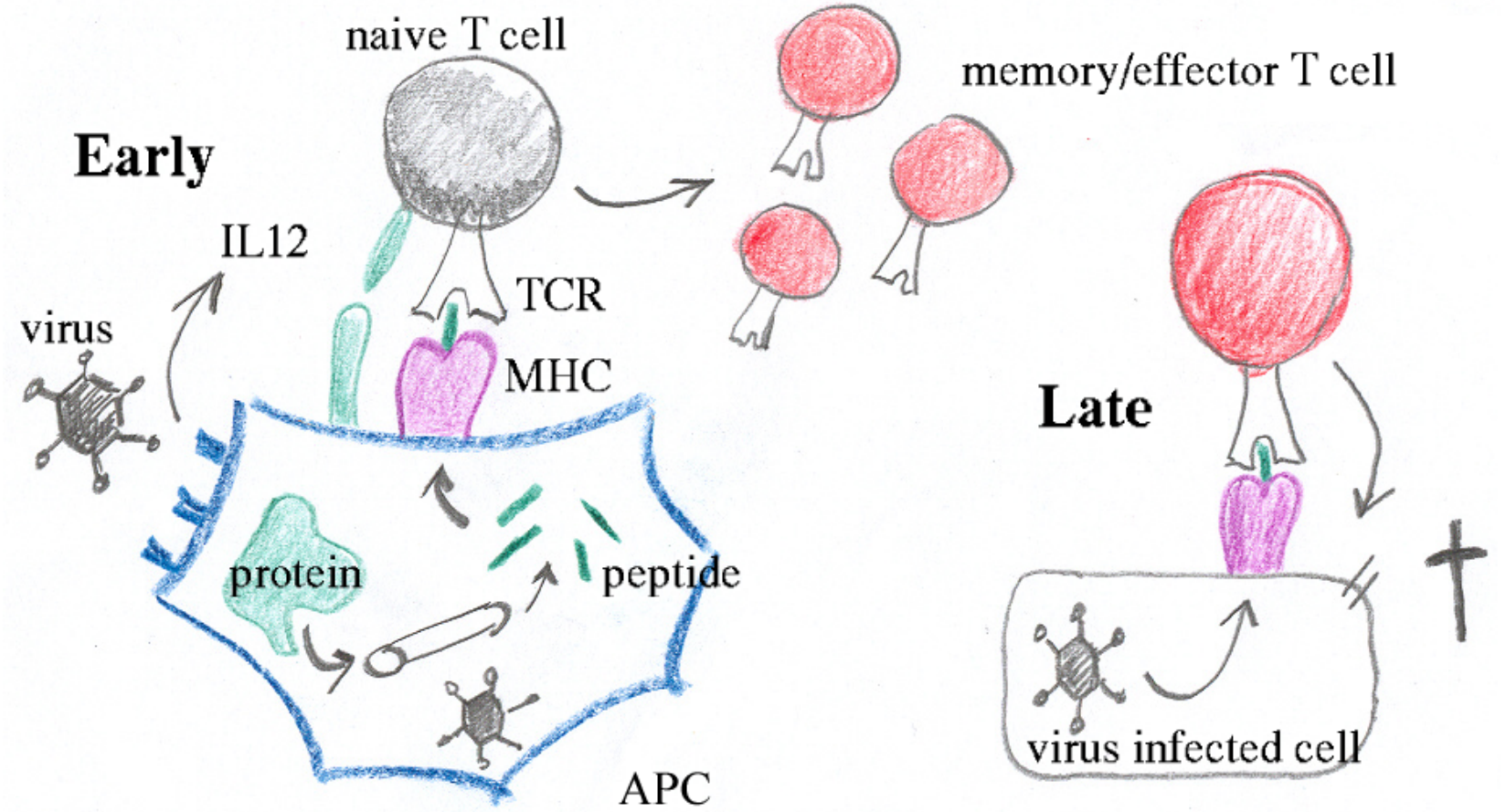
**Information from anchor residues largely preserved**

This works thanks to the specificity of the MHC, and because T cells are restricted to typically one MHC.

## Conclusion: peptides

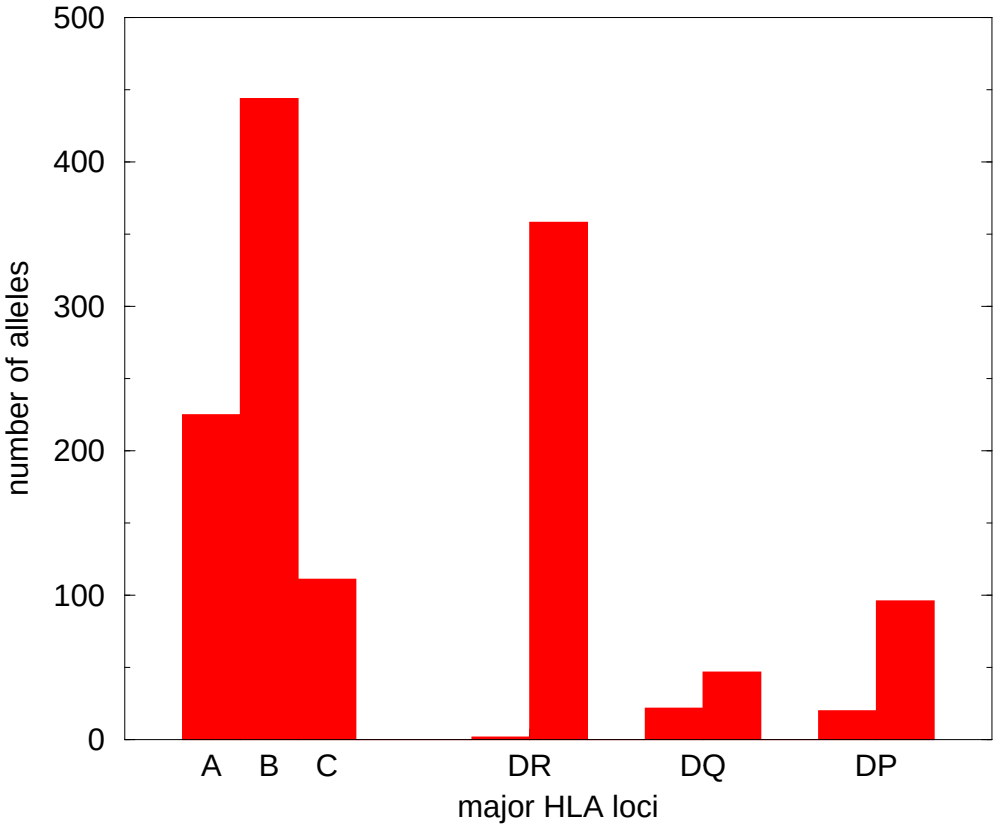
- 9-mers contain sufficient info for self:non-self discrimation
$\rightarrow$ 9-mer overlaps $< 1\%$
- on each MHC several viruses only ten presented foreign peptides among $10^5$ presented self peptides
- MHC specificity sufficient to preserve info from anchor residues
- MHC restriction reduces the overlap between self and foreign and reduces number of self peptides per T cell specificity

# MHC polymorphism




Because MHC is polymorphic we all draw different samples

# Huge HLA polymorphism




From: HLA sequence database.

## Simulating Host-pathogen coevolution

One host population with many pathogen species with MHC molecules and peptides that are both evolving (genetic algorithm).

1. Pathogens random, all hosts the same MHC molecule
2. Infection: every host interacts with every pathogen

$\rightarrow$ yields fitness of hosts and pathogens

3. Replace hosts & pathogens:

$\rightarrow$ fitness proportional selection

$\rightarrow$ fixed population size

4. Allow for mutation during reproduction $\mu_{\text{path}} \gg \mu_{\text{hosts}}$
5. Iterate, i.e., Goto 2.

## Simulate MHC presentation by bitstring matching

- When a pathogen infects a host, check all peptides on all MHC alleles of the host.

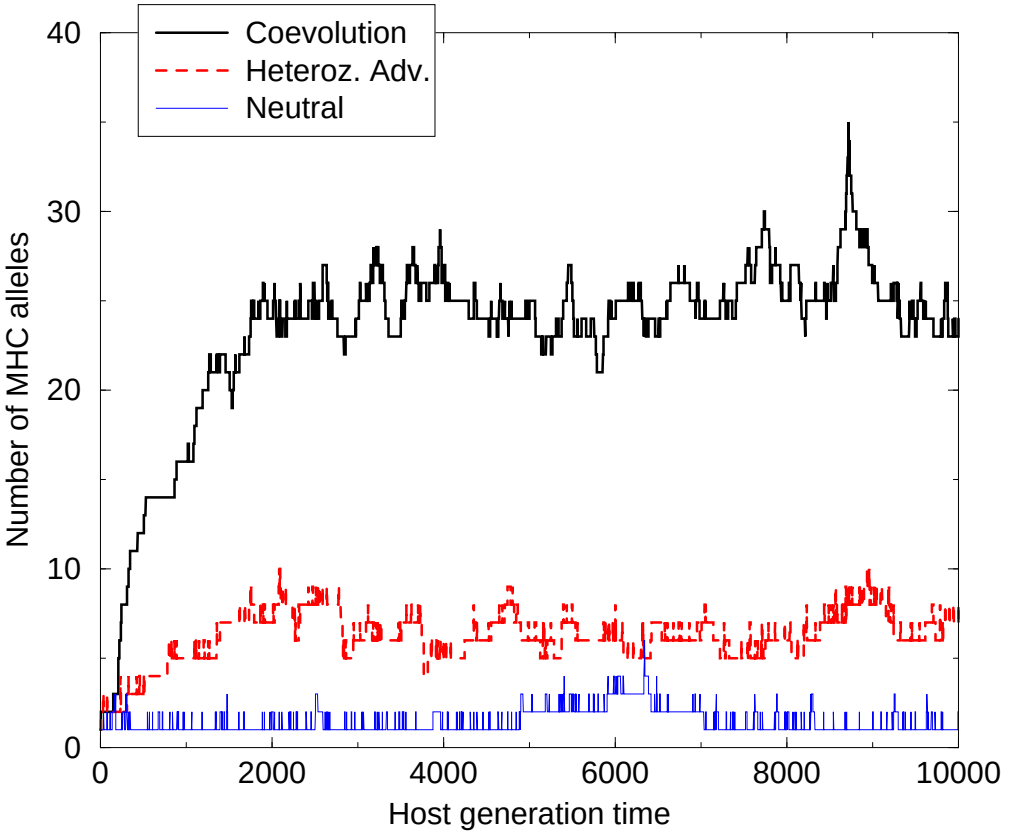
```
MHC:        1010010101011100
peptide:    0110101010100110
match:        **  *********  *
```

- Peptide is presented when longest adjacent match $\geq 7$ bits

$\rightarrow$ Probability that a peptide matches an MHC is 5%

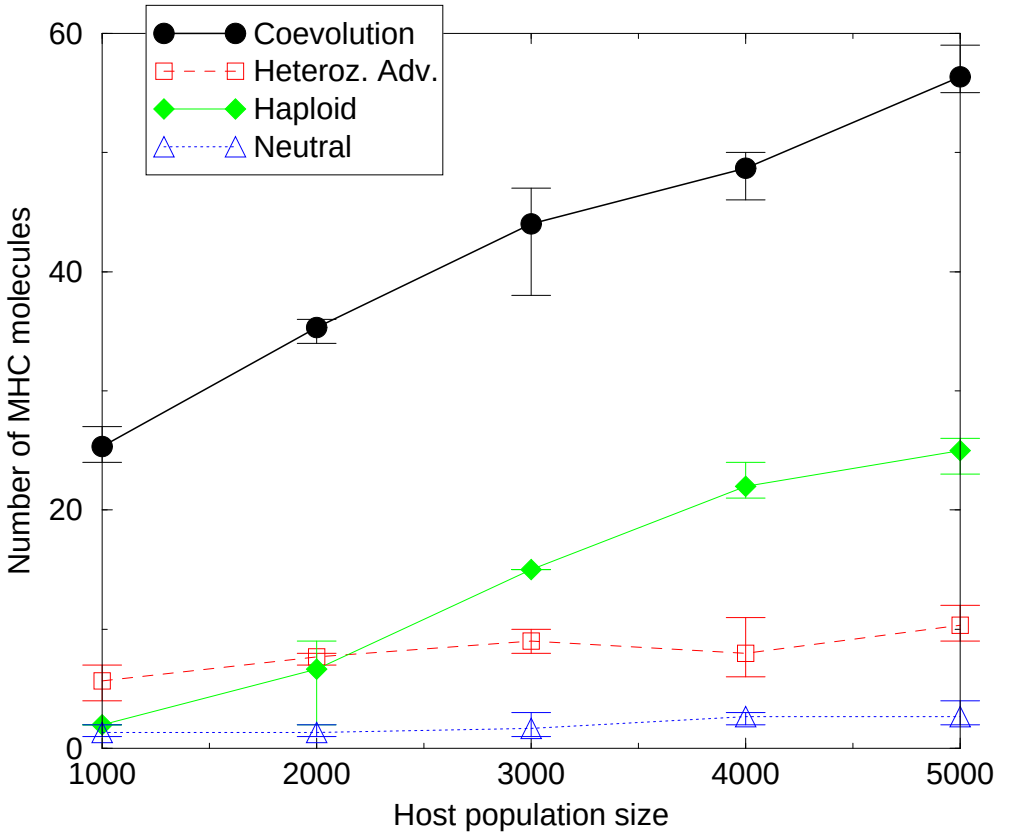
$\rightarrow$ Each host samples a different subset

# Borghans et al., Immunogenetics, 2004




→ Much higher polymorphism with coevolution.

# Polymorphism increases with host population size

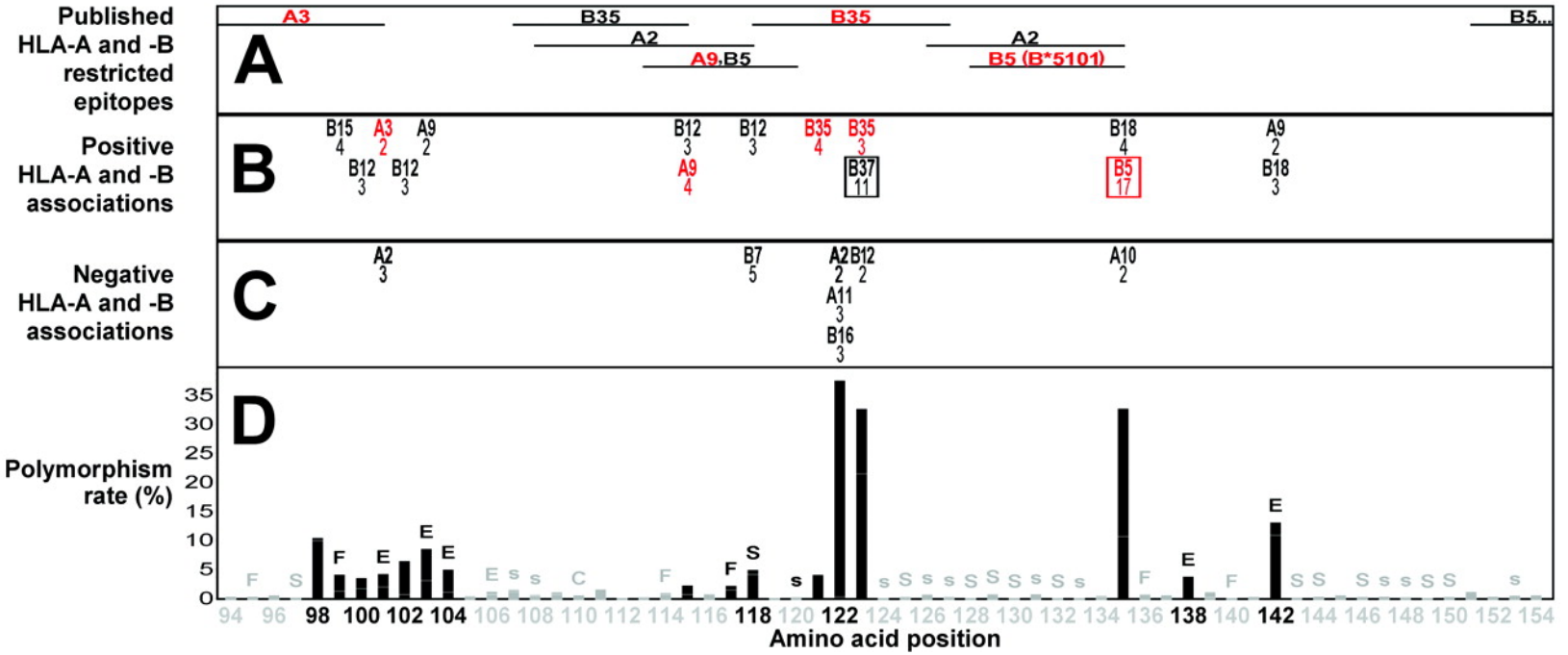



Extinction diminished by increasing population size:

→ many alleles in evolution scenarios

→ few alleles with heterozygote advantage

# Moore et al Science, 2002: HIV-1




Mutations in RT associated with HLA alleles:
evidence for evolution to avoid detection

## De Boer et al., Immunogenetics, 2004

Population genetical model: explaining a high degree of polymorphism by heterozygote advantage requires that the fitness of the alleles is very similar.

To explain a polymorphism of 20 alleles the $20^{th}$ allele should have fitness exceeding 95% of the harmonic mean fitness

This is not to say that there is no heterozygote advantage (HIV, Carrington) but only argues that one would not expect a high degree of polymorphism if heterozygote advantage were the only selection pressure.

## Conclusions

- Host-pathogen coevolution accounts for MHC polymorphism
- 9-mers are sufficiently unique to classify their "source"
- Lymphocyte specificity related to self peptide diversity

### Further questions

- Why is MHC not more diverse (Borghans, Eur. J. Immunol., 2003)?
- Is the non-polymorphic TAP or proteasome the Achilles heel of the polymorphic MHC peptide presentation (Yusim, Keşmir, . . . , Korber, J. Virol., 2002)?

## Collaborators

**Peptide diversity:**

- Can Keşmir (Utrecht University)
- Nigel Burroughs (University of Warwick)

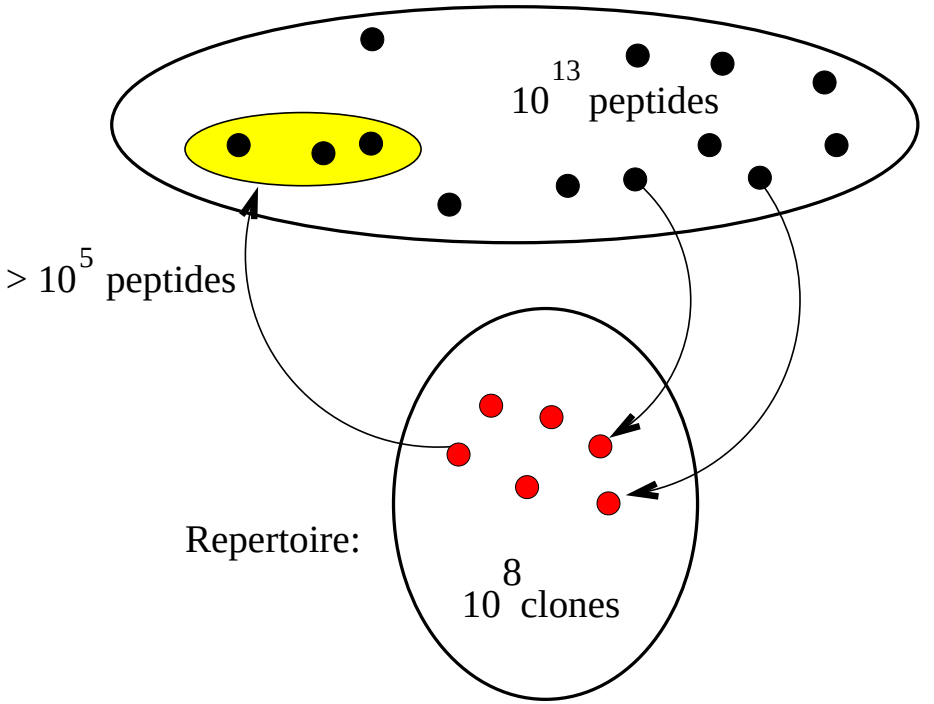
**MHC polymorphism:**

- José Borghans (Sanquin, CLB, Amsterdam)
- Joost Beltman (ITB, Leiden)

**Population Genetical model:**

- José Borghans (Sanquin, CLB, Amsterdam)
- Can Keşmir (Utrecht University)
- Michiel van Boven (Wageningen University, Lelystad)
- Franjo Weissing (University of Groningen)

# Lymphocytes are specific and recognize many peptides




Because there are many more peptides ($20^{10} \simeq 10^{13}$) than clones we need sufficient cross-reactivity (Mason.it98).

**But a peptide recognized by very few clones:** $p \simeq 10^{-5}$

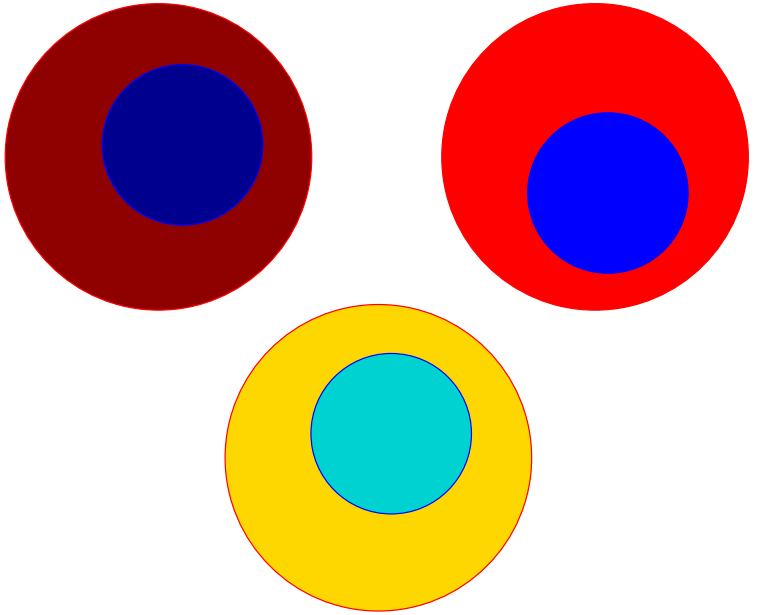
## Strong selection in the thymus

- in mice 5% of the T cells produced in the thymus end up in mature repertoire.
- at least 50% of all positively selected T cells are negatively selected (Van Meerwijk.jem97).

$\rightarrow$ 90% of the thymocytes fail to become positively selected by any of the $M$ MHC molecules in the mouse

$\rightarrow$ **positive selection seems strongest bottleneck**

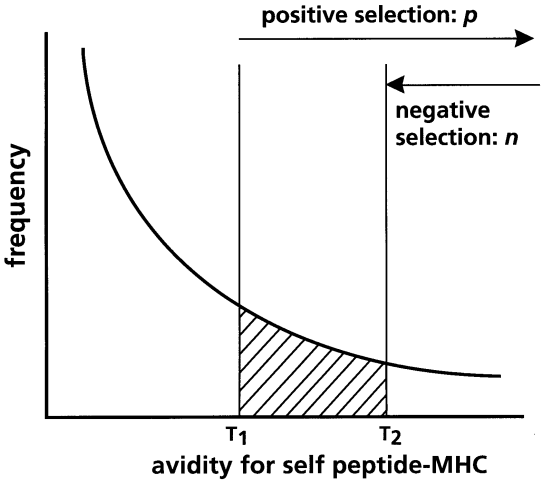
# Additional MHC molecules select largely non-overlapping repertoires



If positive selection is a strong bottleneck, the repertoire size should increase with the number of MHC molecules

$\rightarrow$ we need a quantitative model

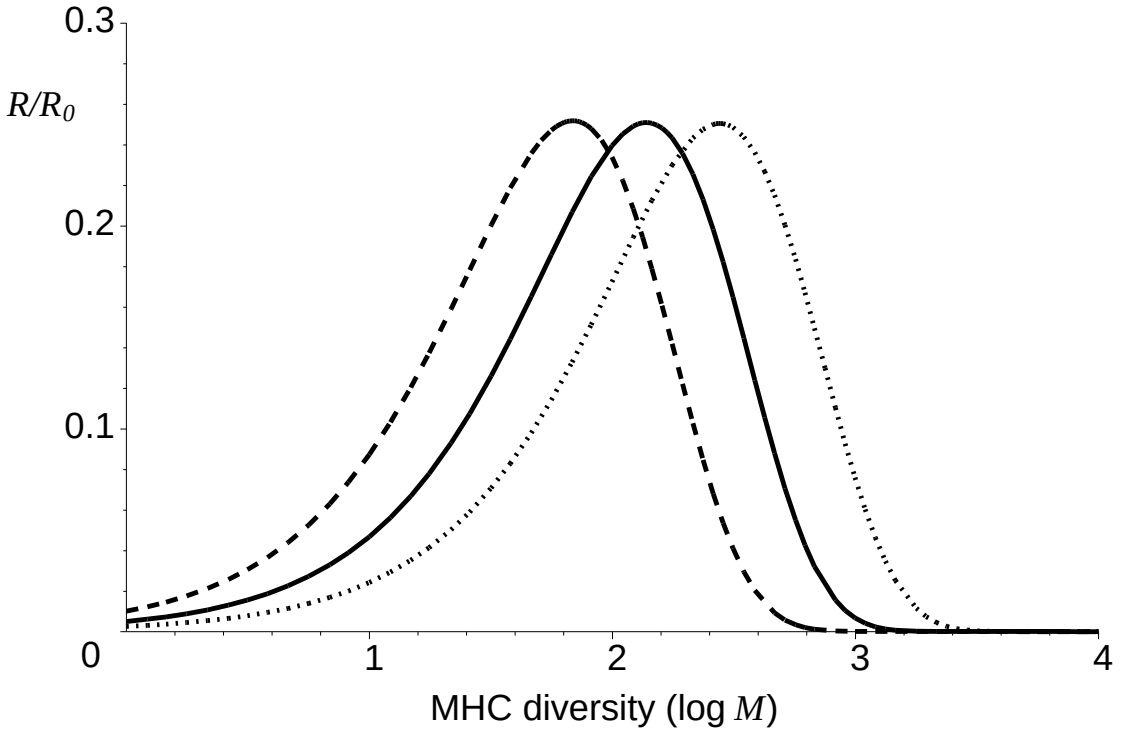
# Size of functional repertoire $R$




$$R = R_0((1-n)^M - (1-p)^M) \quad \text{with} \quad n < p$$

and $M$ is the number of MHC molecules

# Size of the functional repertoire




Solid: $p = 0.01, q = p/2$, dashed: $p = 0.02$, dotted: $p = 0.005$

**For best guess optimum at $M = 140$ MHC molecules**