Identifying selection pressures in somatic immune receptor evolution

Yuval Elhanati[1], Anand Murugan[2], Curtis G. Callan, Jr. [3], Thierry Mora[4] and Aleksandra M. Walczak[1]

***Short Abstract* — Immune receptor diversity enables an effective response to a wide variety of threats. T-cells receptor diversity is the result of initial repertoire generation followed by functional selection of T cells by interactions with self and foreign peptides. Analyzing data on human T-cell receptors, we use maximum likelihood to quantify selection on particular elements of the receptor. We quantify the global and site-specific selection pressures and disentangle selection on amino acids from biases in the generated repertoire. We find correlations between generation and selection of receptors, and a significant reduction of diversity during selection, suggesting natural evolution anticipates somatic evolution.**

***Keywords* — immune receptor diversity, TCR, maximum likelihood, somatic evolution**

I. Introduction

THE immune system is able to recognize and respond to specific pathogenic threats via membrane receptors on T cells. When a T cell receptor (TCR) successfully binds to a pathogenic molecule (antigen), an immune response is initiated. Thus, the effectiveness of the system against a wide range of pathogens depends on a large diversity of different membrane receptors.

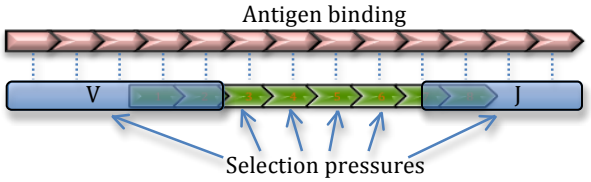
This diversity is initially generated by random combinations of genomic elements - V, D and J genes. The segment in which those genes are joined together, the CDR3 (complementary determining region 3), is highly variable due to random insertions and deletions at the junctions. This segment is known to be central for antigen recognition [1].

The generated sequences are then subject to somatic selection, which results in the functional repertoire. Understanding the role of both generation and selection on repertoire formation, and the relations between them, is key to the understanding of immune diversity.

II. Methods

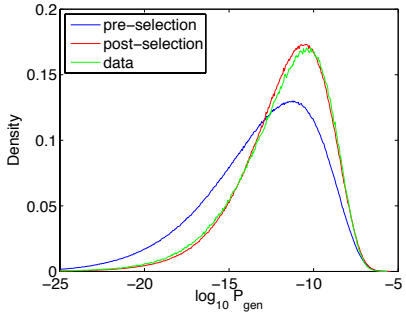
The generation process of the pre-selected repertoire has recently been characterized [2]. We quantify selection pressures on the repertoire by comparing the previously characterized generated pre-selection repertoire to the post-selection repertoire, sampled from data of CD4+ T-cell beta chain DNA sequences [3].

We model selection as depending both on the amino acid content in the CDR3, and on the V and J gene identity, which extend beyond this region. We use maximum likelihood methods to infer the parameters of the selection model.



III. Results

We find that the overall selection pressure on a sequence coding for a particular receptor is well described by independent selection factors. We infer selection factors for both naïve cells, which have been exposed only to self-proteins, and memory cells. We find that the two populations have undergone statistically similar selection processes.



Results indicate that selection significantly limits the diversity of the repertoire, as measured by Shannon's entropy, by eliminating rare receptors. Specifically, generation events that contribute less to the fitness of the receptor are also less likely to occur. This suggests that the random generation process of T cell receptors has evolved to correlate with the somatic selection of the receptors inside the body.

IV. Conclusions

The inferred selection factors for receptors provide a measure of fitness during selection. We find a wide distribution of receptor fitness indicating significant selection even during thymic selection, giving insight into the somatic evolution that shapes repertoire diversity.

References

[1] Murphy KP, Travers P, Walport M, Janeway C (2008) *Janeway's Immunobiology*. Garland, New York.

[2] Murugan A, et al. (2012) Statistical inference of the generation probability of T-cell receptors from sequence repertoires. *PNAS* **109**, 16161-16166.

[3] Robins HS, et al. (2009) Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells. *Blood* **114**, 4099–4107.

[1]Laboratoire de Physique Théorique, CNRS and ENS. Paris, France
[2]Stanford University, Palo Alto, CA
[3]Princeton University, Princeton, NJ
[4]Laboratoire de Physique Statistique, CNRS and ENS, Paris, France