Quantifying the diversity of antibody receptors

Aleksandra M. Walczak

Ecole Normale Superieure, 24 rue Lhomond, Paris 75230 Paris Cedex 05, France

RECOGNITION of pathogens relies on families of proteins showing great diversity. Recent experiments that sequence the entire immune cell repertoires provides a new opportunity for quantitative insight naturally occurring diversity and how it is generated. I will describe probabilistic models of the sequence repertoire, building on a nearly exhaustive sampling of B-cell receptor sequences in zebrafish. Specifically I will focus on describing the diversity in the junctional regions, which is not possible to do by traditional alignment methods. By exploiting the interpretation of these models, I make several predictions for the collective properties of the sequence ensemble: the distribution of sequences obeys Zipf's law (which states that the abundance of sequences decays as the inverse of their rank), the repertoire decomposes into several clusters, and there is a massive restriction of diversity because of the correlations. These results suggest that antibody diversity is not limited by the sequences encoded in the genome and may reflect rapid adaptation to antigenic challenges. I will finish by discussing probabilistic models that describe the generation of diversity in the junctional region in human T-cells that quantify the potential diversity of T-cell repertoires and explain quantitatively why some sequences are shared between individuals. The generative event statistics are highly consistent between individuals, suggesting their origin in a universal biochemical process.