Deciphering the language of fungal pathogen recognition receptors

Witold Dyrka, Pascal Durrens, Mathieu Paoletti, Sven Saupe, David Sherman

DECIPHERING THE LANGUAGE OF FUNGAL PATHOGEN RECOGNITION RECEPTORS

Witold Dyrka[1,2,3]*, Pascal Durrens[1], Mathieu Paoletti[2], Sven J Saupe[2], David J Sherman[1]

[1] *INRIA - Université Bordeaux - CNRS, Team MAGNOME, Talence, France*

[2] *Institut de Biochimie et Génétique Cellulaires, CNRS, Bordeaux, France*

[3] *Wrocław University of Technology, Department of Biomedical Engineering, Poland*

**presenting author, e-mail: witold.dyrka@pwr.edu.pl*

The immune system searches for pathogen invasion markers, which include pathogen proteins and host proteins modified in the course of the invasion. While some pathogen-associated molecular patterns are relatively invariant, numerous pathogen-specific markers change quickly. Therefore, to win the arms race with pathogens, host recognition receptors must adapt quickly to varied and modulating markers. To achieve this goal, the recognition domain of the receptor requires the capacity to recognize diverse possible pathogen molecule epitopes, and ability to quickly learn new ones. In plants and fungi, which lack an adaptive immune system, this key role in the immunity is played by the NLR family of receptors, which adapt to ever-changing pathogen-specific invasion markers thanks to their repeat-based architecture, which can produce diversity of recognition paratopes through unequal crossing-over and mutation [1]. The unequal crossing-over is a repeat shuffling process 10,000 to 100,000 times quicker than the standard point mutation, however it requires (and promotes) high sequence similarity between consecutive repeat units. At the same time certain positions in the repeats are highly variable under positive diversifying selection; they are believed to form the actual recognition paratope [2, 3]. Characterizing computationally the language of these pathogen recognition receptors can provide insight into the molecular mechanisms of immune response and describe the limits of the pathogen targets that can be recognized.

In this work, we modeled generation and selection of the recognition paratopes as a stochastic string rewriting system with constraints, tuned by analysis of observed evolutionary processes and validated with regard to a large data set of fungal NLR [3]. First, we cross-checked the unequal crossing-over and mutation model with the existing experimental data. Second, we analyzed mathematical properties of the model and showed that its convergence to stationary distribution. Then, we compared real and simulated data. Among others, analyzing the feasible set of solutions revealed that the model explained the i=i+2 (even-odd) periodicity observed in the repeat number distribution of a family of receptors. Next, we explored discrepancies between real and simulated data in order to discover constraints acting on the paratopes. For example, by comparing amino acid content in the model and original populations, we confirmed that highly variable sites identified on the basis of entropy, were subject to constraints towards composition typical for binding sites, which was coherent with the suggested role of recognition paratopes. Finally, we proposed an interactive approach to exploring the solution space of amino acid repeats by means of 2-d projections and significance tests in sectors of the space. In a preliminary analysis, we found an overrepresented pattern R-[SYQFW](1,3)-R at one of the highly variable positions in a family of receptors, which potentially has functional importance.

The methodology developed in this work is general and therefore can be applied to any class of amino acid repeats generated by unequal crossing-over for which an equivalent high quality data set is available. An appealing future application of our model would be testing of hypotheses regarding repeat origin or function encoded as constraints on the string rewriting system.

[1] M.Paoletti, S.J.Saupe. *BioEssays*, **2009**, 31:1201-1210

[2] M.Paoletti, S.J.Saupe, C.Clave, PLoS ONE, **2007**, 2:e283

[3] W.Dyrka, M.Lamacchia, P.Durrens, B.Kobe, A.Daskalov, M.Paoletti, D.J.Sherman, S.Saupe, *Gen Biol Evol*, **2014**, 6, 3137-3158.