The evolution of an FLS2 gene responsible for pathogen sensing in Arabidopsis

The purpose of this project is to determine the level of purifying and positive selection in the Arabidopsis gene FLS2 which is a pattern-recognition receptor responsible for the signaling cascade following a pathogen detection. The FLSC2 gene is important due to its involvement in pathogen flagella sensing on the extracellular matrix of a cell where it is the first promoter of the signaling cascade which induces a defensive response (Koller and Bent 2014). This particular gene, however, is also found to be in mammals and insects alike which shows its pervasiveness and importance in biological lifeforms (Gomez-Gomez and Boller 2011). It is therefore one of the main lines of pathogen sensing for cells, and it is responsible for the initial detection of pathogen presence. The gene also plays another role besides detection; because it is also the main driver of the signaling cascade involving the pathogen-associated molecular patterns (PAMP), along with being the initial activator of the innate immune MAP kinase signaling cascade, both of which have varying levels of phosphorylation and defensive gene regulation in plants (Koller and Bent 2014). Due to its importance in pathogen sensing, the gene must have some type of selection imposed on it and it may be subject to neutral mutations.

Seeing as the FLS2 gene encodes a protein that is responsible for sensing flagellum from bacteria it is not a stretch to say that there may be some sort of selection to increase the ability of the protein to bind to the MAP kinase along with its involvement in BAK1 (Koller and Bent 2014). This selection could cause the activity rate of the FLS2 gene to increase, or for the protein itself to bind and transfer phosphate groups at a higher rate than before. If there is a lack of positive selection for the gene of interest, there also could be neutral mutations amassing at high levels in the protein which do not detrimentally affect the conformation and function of the protein or give a fitness advantage/disadvantage. However since this particular gene is found throughout biological life it is also very likely that the gene itself is under purifying selection in order to maintain the function of flagellin sensing since most bacteria have a flagellin, or a flagellin like, structure used for motility. In order to test these hypotheses multiple levels of analysis will be used in order to discern the characteristics

In order to conduct this research I will gather sequences of the FLS2 gene from the 1001 genomes database - where there are genomic sequences involving 19 arabidopsis accessions available for download. By using these 19 accessions, I will have a better view of the protein sequence evolution over ecological space. The first step would be to align the sequences using a program like SeaView or SAMTools, and then to use the CodeML package in order to compute both the kappa and dN/dS values for the acessions. The next step would be to integrate the coding data into a phylogenetic tree in order to see the relationship between the sequences -- if there is indeed differentiation in coding sequences. If using the 19 accessions is not enough for a tree to be built reliably, then the SNP data from the 1001 genomes project could be used to infer another tree using more accessions from the 19 populations alongside calculating Tajima’s D if the SNP data are used alongside a Col-0 genotype to be used as the base sequence for comparison. At that point the data should give indications of polymorphic sites in the sequences which could then be loaded up into Pymol in order to visualize the polymorphic sites on the actual protein structure which would allow for a qualitative visualization of the residues in the structure. Using this methodology one would be able to discern what type of selection and constraints are being placed on this gene at the population level.

Justin: This is an interesting system and you’ve got a toehold on some good analyses. The logic of the proposal seems unduly porous. The flagellin-detecting system would seem to be an excellent case in which a host-pathogen arms race produces the classic signature of positive selection, rapid amino acid replacement quantifiable by dN/dS, at the site of interaction with the pathogen. The rest of the protein would, as you say, be under purifying selection to maintain interaction with the MAP kinase downstream. Quantifying the signatures of selection here, and mapping these sequence-derived signatures onto a structure for FLS2, would be a strong proposal. It seems as though you’re leaning that direction, but the clear statement of what you plan to do isn’t quite there. Hope this is helpful; please feel free to seek me out if discussion would be useful. -- Allan

References

Koller, Teresa, and Andrew F. Bent. "FLS2-BAK1 extracellular domain interaction sites required for defense signaling activation." *PLoS One* 9, no. 10 (2014): e111185.

Gómez-Gómez, Lourdes, and Thomas Boller. "FLS2: an LRR receptor–like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis." *Molecular cell* 5, no. 6 (2000): 1003-1011.