

INTERPRETATION OF RESULT

Khoa Thanh Nguyen, Huong Phan

We picked *Folsomia candida* as our Collembola organism of choice. When we BLAST the protein fasta of the fruit fly (*Drosophila melanogaster*) Osiris protein into the proteome database of *Folsomia candida*, we were able to see matches in all 24 Osiris genes. Then, we only consider matches with E-value less than 0.05, which is seen in Osi-17 (FBpp0078240), Osi-24 (FBpp0078259) and Osi-15 (FBpp0078256), with Osi-15's best hit having much higher E-value than that of Osi-17 and Osi-24. According to Smith et al. 2018, these two Osiris genes belong to the similar "blue" clade and thus belong to a cluster. With other matches for Osiris genes, the E-value is more than 0.05 so the result is not significant enough to conclude that we find evidences of those Osiris genes in *Folsomia candida*. The best hit is the ID XP_021967935.1, corresponding to query for Osi-24.

query_id	subject_id	pct_iden tity	aln_leng th	n_of_mis matches	gap_op enings	q_sta rt	q_en d	s_star t	s_end	e_value	bit_score
FBpp0078240	XP_021967935.1	26.667	105	67	4	404	505	110	207	5.27E-07	50.8
FBpp0078240	XP_021961172.1	28.571	91	61	2	417	505	112	200	0.001	40.4
FBpp0078240	XP_021961171.1	28.571	91	61	2	417	505	113	201	0.002	40.4
FBpp0078256	XP_021965938.1	29.487	78	51	1	36	109	532	609	0.043	34.7
FBpp0078259	XP_021967935.1	27.778	108	63	4	550	647	117	219	9.49E-07	50.4

We retrieve the protein sequence for that ID XP_021967935.1 and BLAST it back into the protein fasta of *Drosophila melanogaster* database. Upon observing the result,

we again see the E-value less than 0.05 for Osi17 and Osi24, with Osi24 the one with lowest E-value (5.27E-07), and all of those matches again belong to XP_021967935.1. Since we find the same gene as best hit in our rbhBLAST for Osiris proteins with significant E-value (1.45E-10), we can conclude that we find 1:1 orthologs for XP_021967935.1. For other Osiris genes, we are not certain because the result is not significant enough as E-value is greater than 0.05. Since we're able to see the evidence of Osiris genes in *Folsomia candida*, a Collembola species, the Osiris cluster is not unique to insects.

query_id	subject_id	pct_iden tity	aln_len gth	n_of_mis matches	gap_op enings	q_st art	q_en d	s_sta rt	s_end	e_value	bit_score
XP_021967935.1	FBpp0078240	26.667	105	67	4	110	207	404	505	1.45E-10	50.8
XP_021967935.1	FBpp0078259	27.731	119	70	5	107	219	539	647	2.7E-10	50.1

From looking at the Figure 1. in Smith et al. 2018, we hypothesize that Osi24 genes arise fairly early compared to other Osiris genes, during the time insects and non-insects diverged from each other. Osi17 arises from the same clade as Osi24, which we think might be a gene duplication event. Since these events happen during the time insects diverge from Collembola, it makes sense that there can be evidences of Osi24 and Osi17 in Collembola.

Citations:

Smith, C. R., Morandin, C., Nouredine, M., & Pant, S. (2018). Conserved roles of Osiris genes in insect development, polymorphism and protection. *Journal of evolutionary biology*, 31(4), 516-529.