

# DDM

# ANN and Metaproteomics Update

Victor Seguritan

April 19, 2012



SAN DIEGO STATE  
UNIVERSITY

# Manuscripts

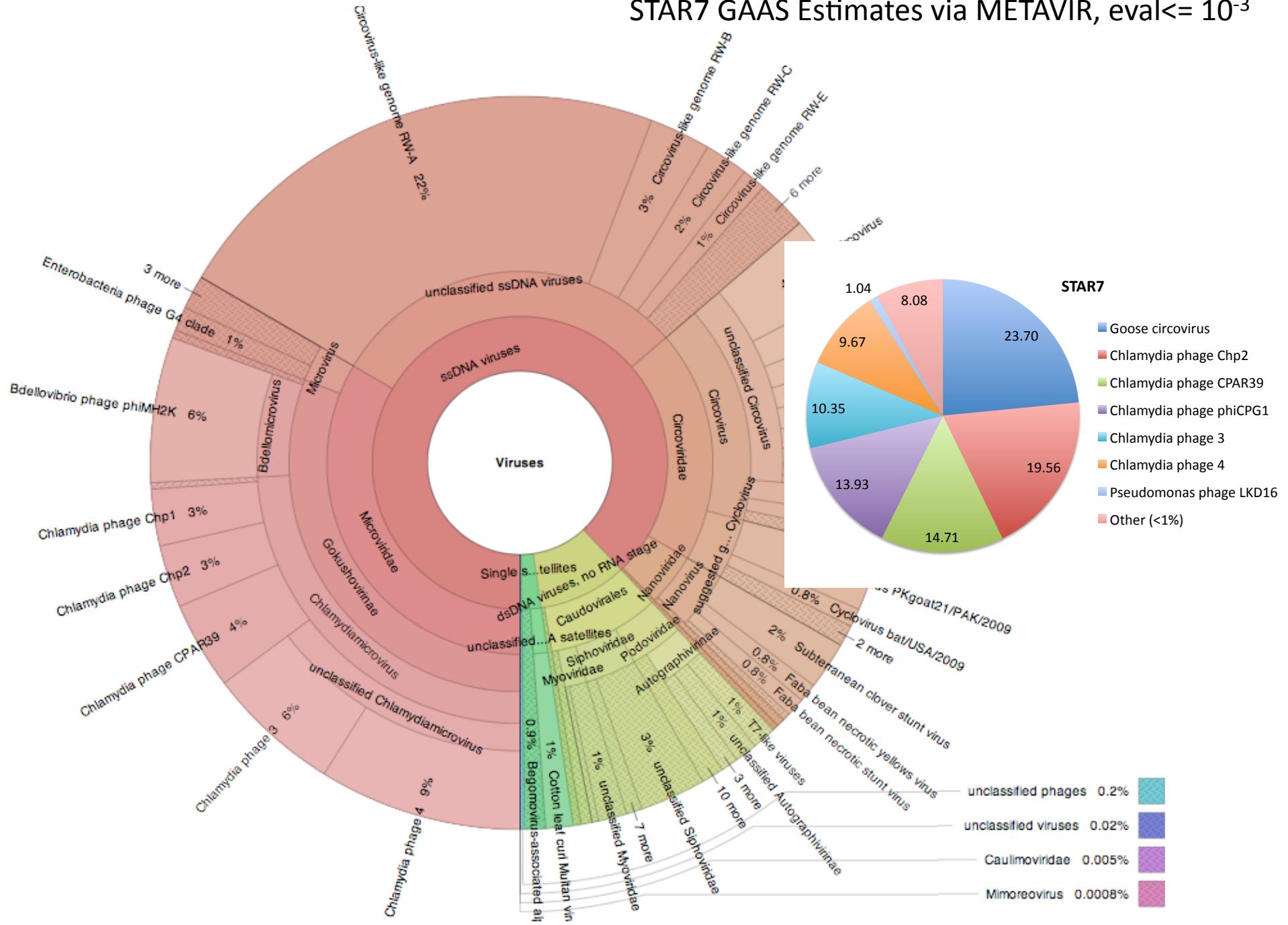
## Structural Protein Neural Networks

- Manuscript was resubmitted and under review.

## Metaproteomics

- GAAS Results
- MASCOT / MOWSE Searches
- Scripts for automation

# STAR7 GAAS Estimates via METAVIR, eval<= 10<sup>-3</sup>

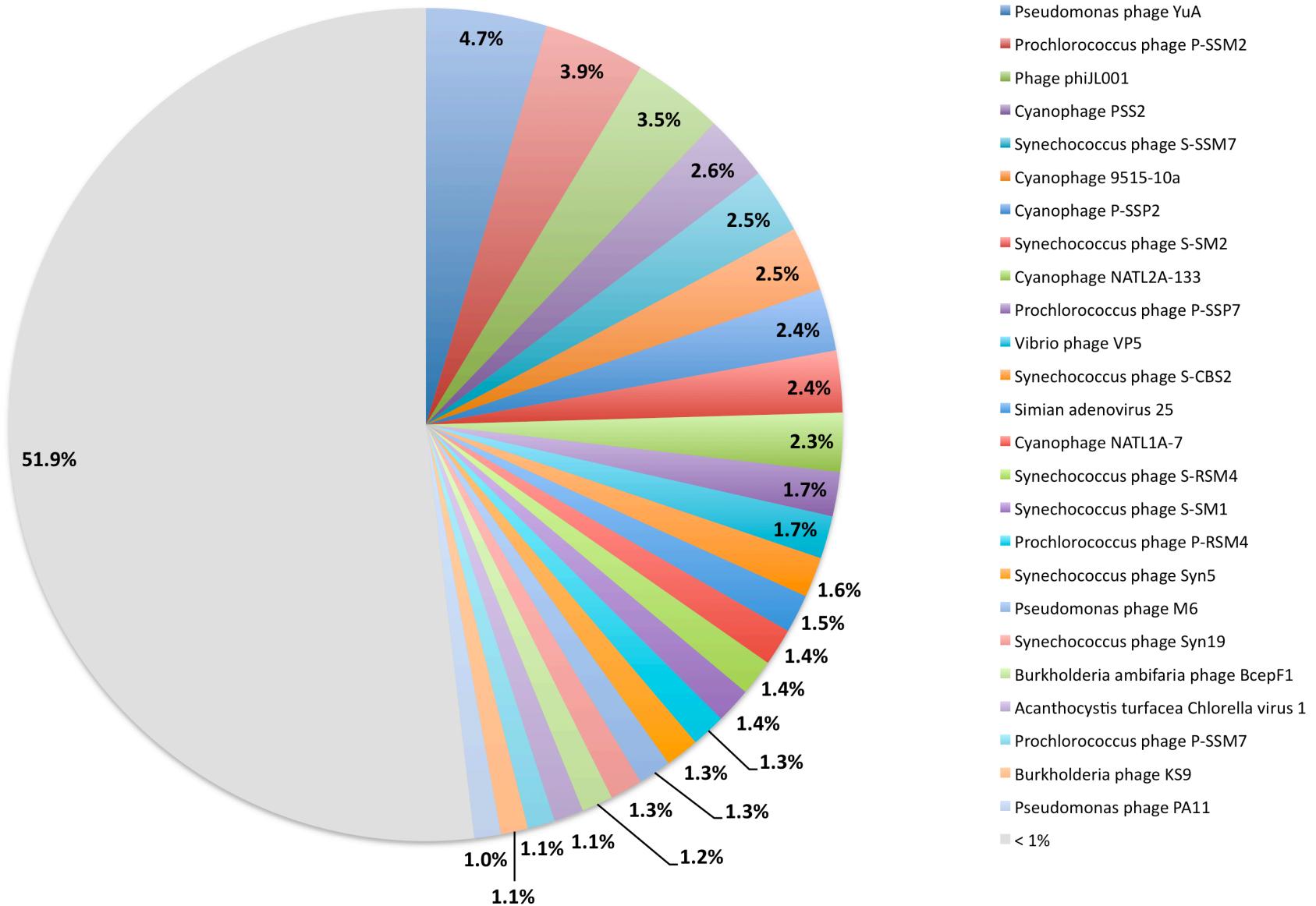


STAR7

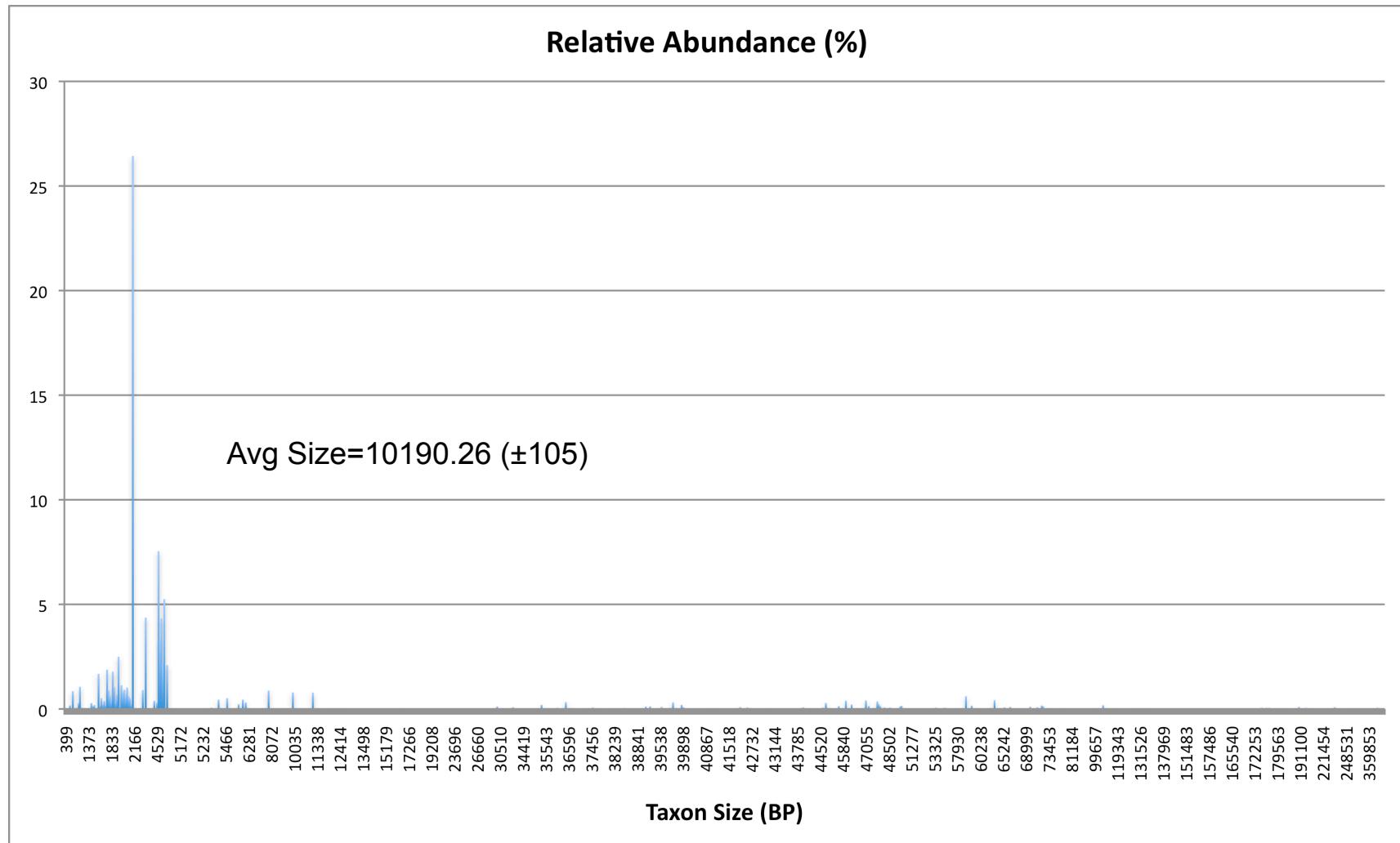
Goose circovirus	23.70
Chlamydia phage Chp2	19.56
Chlamydia phage CPAR39	14.71
Chlamydia phage phiCPG1	13.93
Chlamydia phage 3	10.35
Chlamydia phage 4	9.67
Pseudomonas phage LKD16	8.08
Other (<1%)	1.04

unclassified phages	0.2%
unclassified viruses	0.02%
Caulimoviridae	0.005%
Mimoreovirus	0.0008%

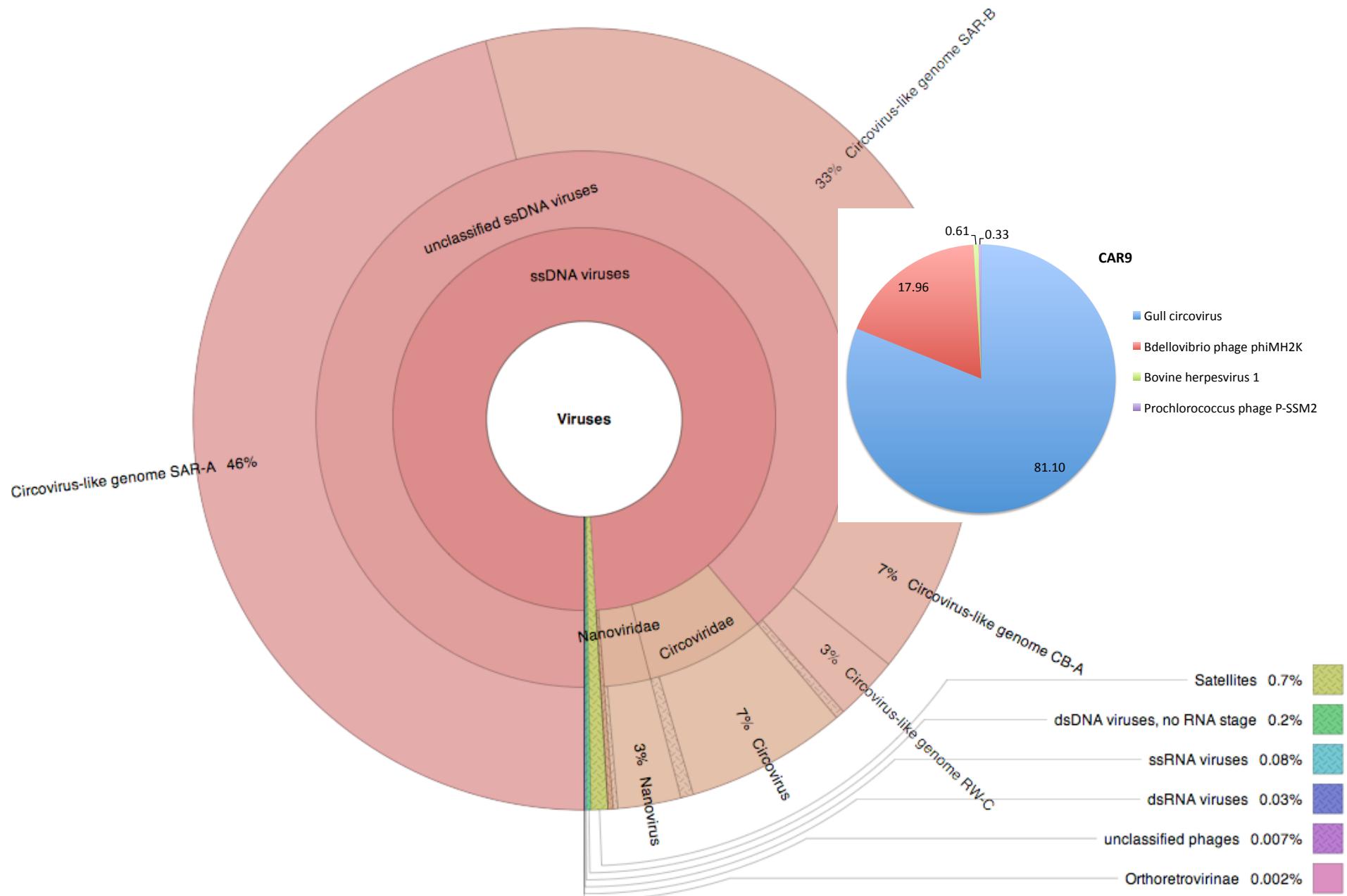
# STAR7 GAAS Estimates (minus Micro-, Nano-, and Circoviridae)



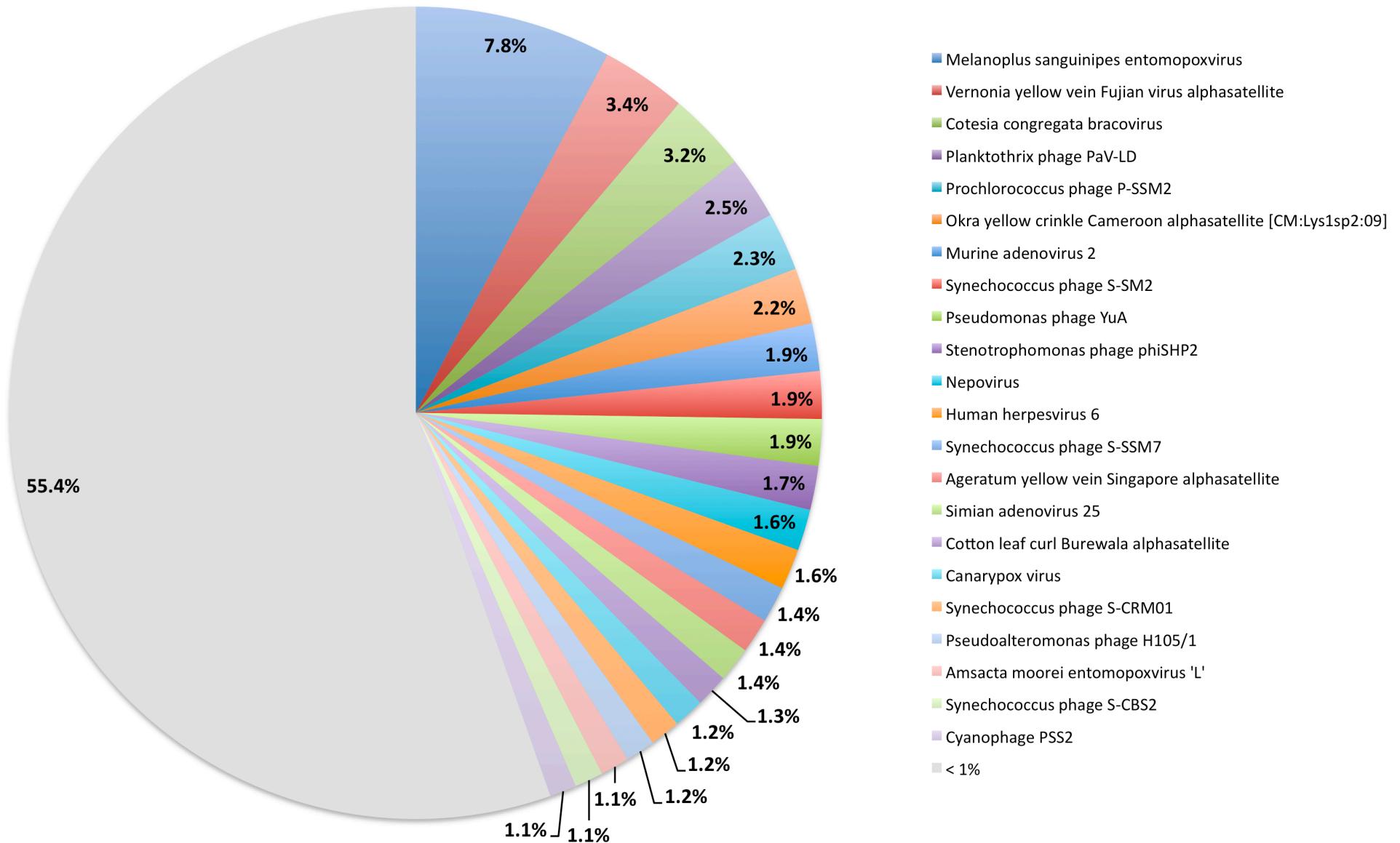
# STAR7 Taxon Sizes



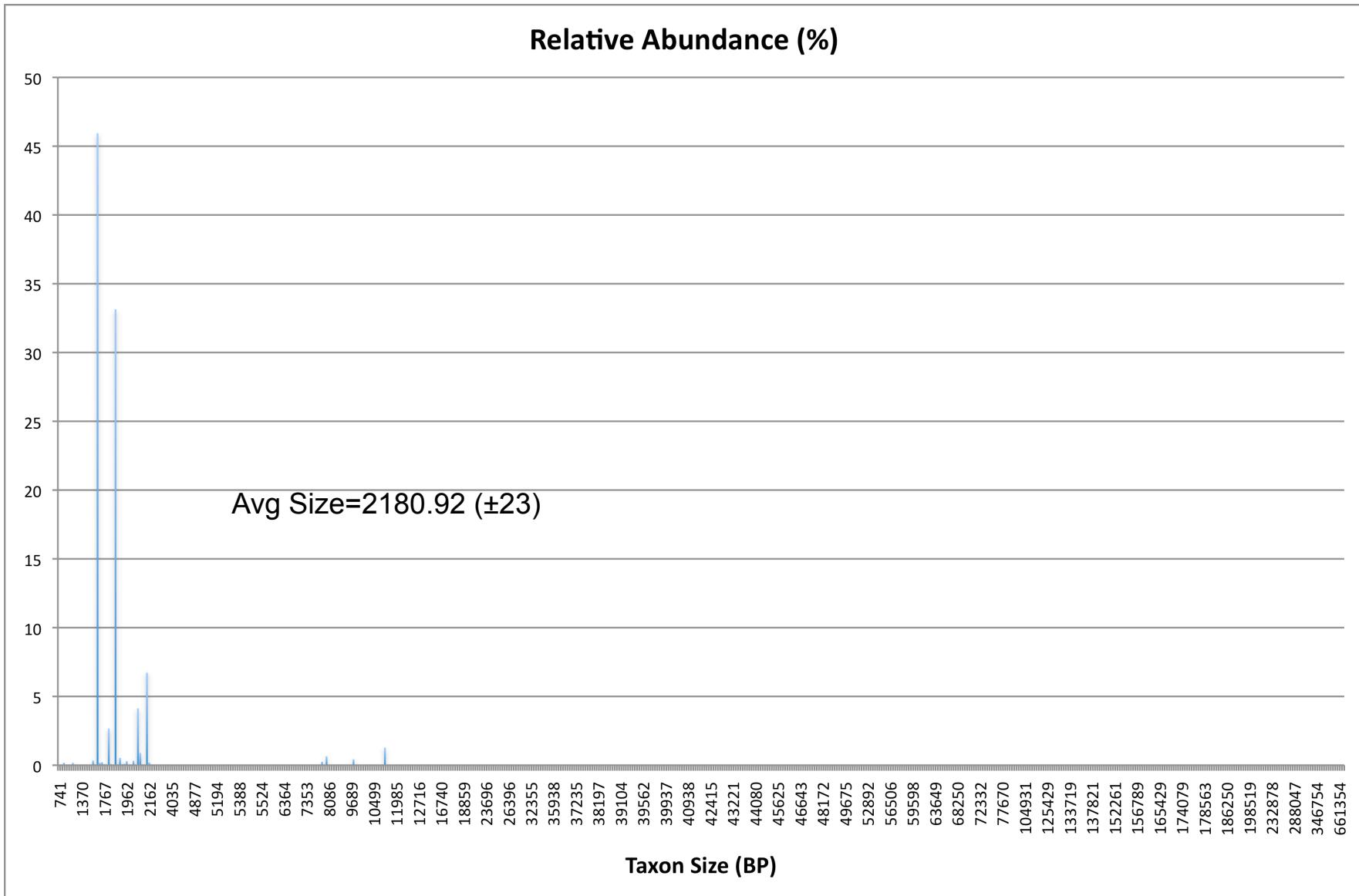
## CAR9 GAAS Estimates via METAVIR – eval<= 10<sup>-3</sup>



# CAR9 GAAS Estimates (minus Micro-, Nano-, and Circoviridae)

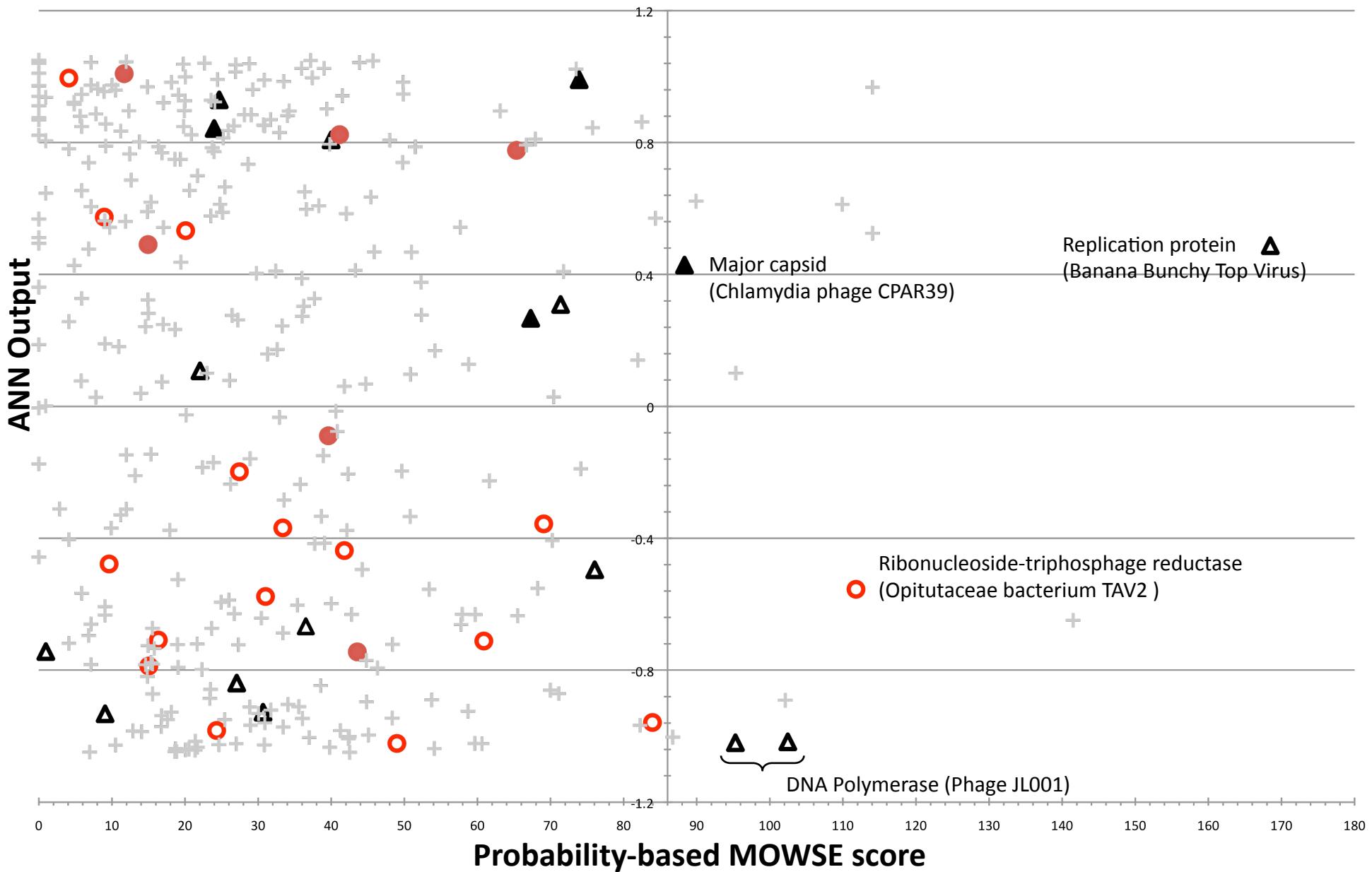


# CAR9 Taxon Sizes



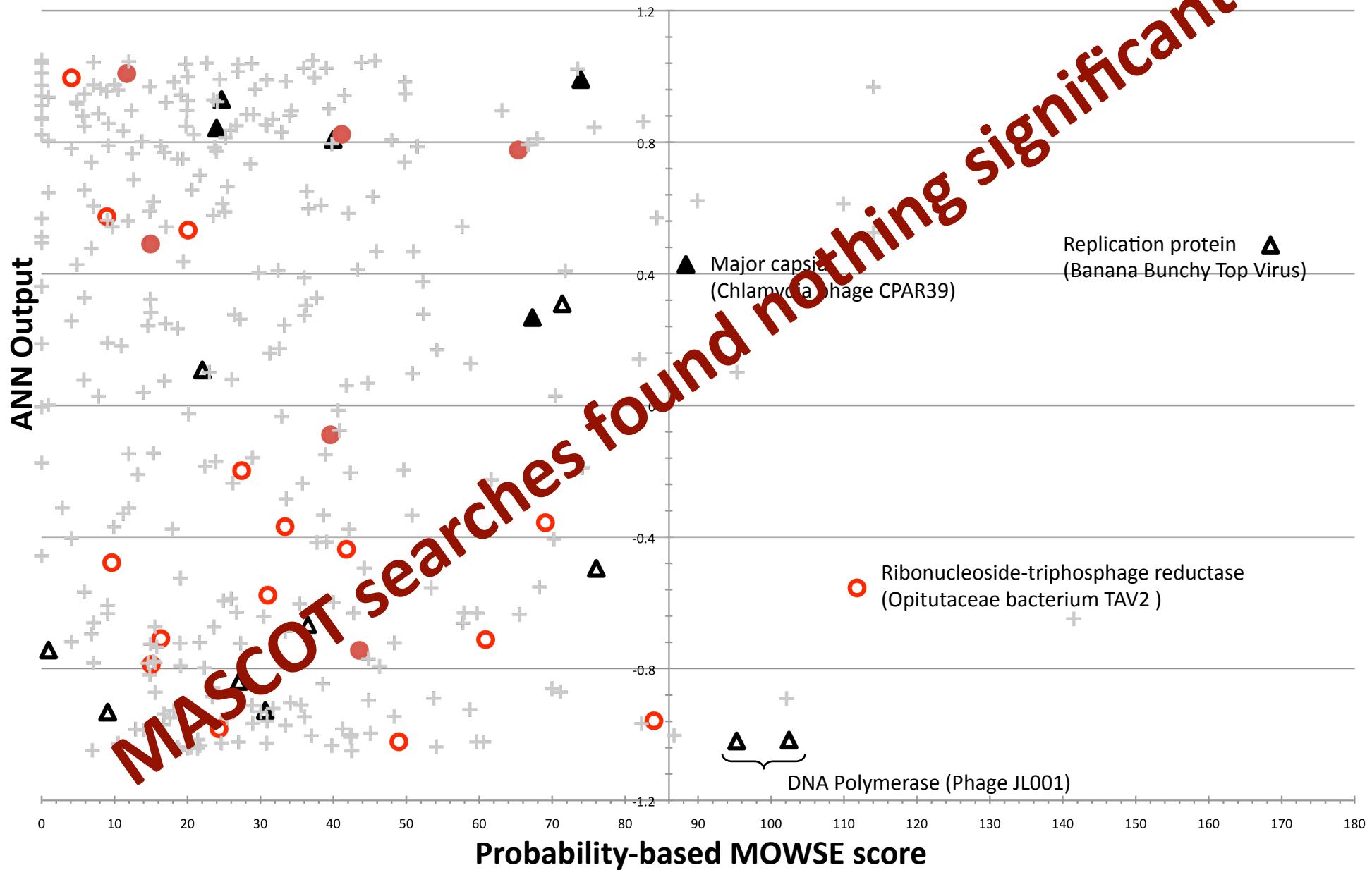
# Protein Identification by MOWSE

▲ Virus Structural   ▲ Virus Non-structural   ● Pro-phage Structural   ○ Prokaryote Non-structural   + Unknowns



# Protein Identification by MOWSE

▲ Virus Structural   ▲ Virus Non-structural   ● Pro-phage Structural   ○ Prokaryote Non-structural   + Unknowns



# MASCOT Search Results - NR

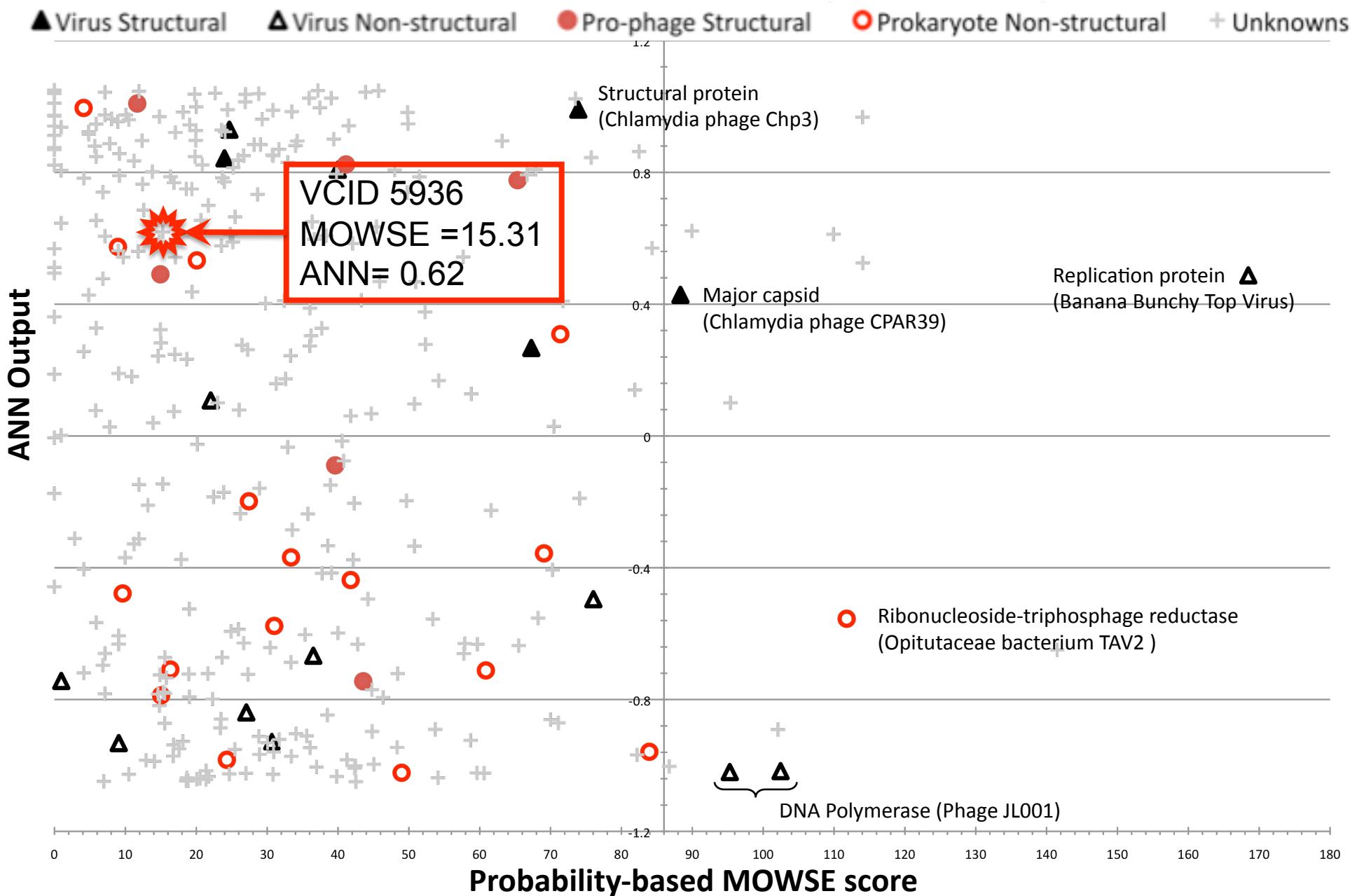
MS data file : AnalysisStar7.mgf  
Database : NCBIInr 20120421 (17910093 sequences; 6147033692 residues)  
Timestamp : 23 Apr 2012 at 11:25:01 GMT  
Protein hits :  
[gi|11935049](#) keratin 1 [Homo sapiens]  
[gi|623409](#) keratin 10 [Homo sapiens]  
[gi|61740600](#) keratin, type I cytoskeletal 10 [Canis lupus familiaris]  
[gi|146741296](#) keratin 1 [Sus scrofa]  
[gi|351695070](#) Keratin, type II cytoskeletal 1, partial [Heterocephalus glaber]  
[gi|4159806](#) type II keratin subunit protein [Mus musculus]  
[gi|12859782](#) unnamed protein product [Mus musculus]  
[gi|136429](#) RecName: Full=Trypsin; Flags: Precursor  
[gi|348562686](#) PREDICTED: keratin, type I cytoskeletal 10-like [Cavia porcellus]  
[gi|348581121](#) PREDICTED: keratin, type II cytoskeletal 1-like [Cavia porcellus]  
[gi|435476](#) cytokeratin 9 [Homo sapiens]  
[gi|194037336](#) PREDICTED: keratin, type II cytoskeletal 1b [Sus scrofa]  
[gi|148727309](#) keratin, type II cytoskeletal 2 epidermal [Pan troglodytes]  
[gi|109099462](#) PREDICTED: keratin, type II cytoskeletal 6B-like [Macaca mulatta]  
[gi|149031970](#) rCG50690 [Rattus norvegicus]  
[gi|355698804](#) keratin 6A [Mustela putorius furo]  
[gi|73996461](#) PREDICTED: keratin, type II cytoskeletal 78 [Canis lupus familiaris]  
[gi|291334540](#) prophage LambdaCh01 coat protein [uncultured phage MedDCM-OCT-S04-C1220]  
[gi|52789](#) unnamed protein product [Mus musculus]  
[gi|293686](#)  
  
MS data file : AnalysisCar9.mgf  
Database : NCBIInr 20120421 (17910093 sequences; 6147033692 residues)  
Timestamp : 23 Apr 2012 at 11:35:03 GMT  
Protein hits :  
[gi|160961491](#) keratin, type II cytoskeletal 1 [Pan troglodytes]  
[gi|291389217](#) PREDICTED: keratin 6A-like [Oryctolagus cuniculus]  
[gi|348581121](#) PREDICTED: keratin, type II cytoskeletal 1-like [Cavia porcellus]  
[gi|4159806](#) type II keratin subunit protein [Mus musculus]  
[gi|12859782](#) unnamed protein product [Mus musculus]  
[gi|291389225](#) PREDICTED: keratin 8-like [Oryctolagus cuniculus]  
[gi|136429](#) RecName: Full=Trypsin; Flags: Precursor  
[gi|47604942](#) keratin 75 [Gallus gallus]  
[gi|224099135](#) PREDICTED: similar to cytokeratin type II [Taeniopygia guttata]

No significant hits against STAR7 pORFs

# MASCOT Hits to STAR7 Sequences

Seq ID	BLASTP	ANN	MOWSE	E-val	MASCOT Hits			
					Peptide 1		Peptide2	
					Sequence	E-val	Sequence	E-val
16031_6_2	hypothetical protein	-0.67	15.5518	2.1	ISEQAAVQMPMKTVASLIAMIAVGTWAYFGIHEK			
2146_6_2	hypothetical protein	0.97	7.10127	3.7	TEYDGSTWTTKSNSMGVSLYR			
26232_3_1	No hits found	0.59	25.1388	3.6	QIEKLFK	3	NQLRQYGLPAFANGGIVGMGPGQSR	
4336_5_15	hypothetical protein	-0.33	11.2065	3.2	SSLVVGKR			
27756_6_69	hypothetical protein	0.27	36.0452	3.1	VQASIPVR			
28154_4_11	ribonucleoside-triphosphate reductase [Roseiflexus castenholzii DSM 13941]	0.31	71.3758	3	GKPEMGTMR			
24501_1_5	hypothetical protein	0.28	52.3631	3	LEVQLALPCR			
6482_3_19	structural protein [Chlamydia phage Chp2]	0.99	73.9156	2.6	LQDPEYLGGGSNR			
10190_1_1	hypothetical protein	-0.59	25.9649	2.3	EGNQQYALVDMKNE	1.9	GENPPEPQTEPLVTMFTK	
4335_3_26	No hits found	0.10	50.8172	2.1	QRLLPTMMQR			
24005_3_7	hypothetical protein	0.93	23.5737	2.1	SRAQFGDSMWTGAR			
2554_5_29	hypothetical protein	1.03	35.9551	1.8	DPDMGIMRPPPGMSRLPK			
4461_3_3	hypothetical protein	0.85	30.8636	1.7	KAAEER	5.1	KAAEER	
13476_5_2	No hits found	-0.63	9.06241	1.7	LGMSVYR			
2565_3_14	hypothetical protein	0.30	36.2046	1.4	GLNVGFLAK			
21676_2_21	hypothetical protein	-0.66	57.755	1.1	VQELLKDEMIHRPR			
2890_4_64	hypothetical protein	1.05	37.1753	0.95	MALIPVTTPAGIVKNGTEYATK			
9341_2_6	hypothetical protein	-0.20	49.6547	0.61	EYVDTLNYNPPQTSSIEDNK			
11954_2_1	No hits found	1.04	26.9663	0.49	VSAKPGTNFTIWDTGERGELK			

# ANN vs MOWSE (STAR7)



# Summary

## Metagenome Sequence Analysis

- GAAS shows that small genomes are still prevalent.

## Mass Spec. Fingerprinting

- No dominant species of peptides
- Protein concentration too low
- Structural protein sequences and 5936 were not detected by MS

# In Progress

- Metaproteomics Paper
  - Include METAVIR (GAAS, MDS, etc.) analysis with caveat, i.e. phi29 sequencing bias
  - Emphasize matches between the proteome and MG sequences (downplay MASCOT searches)
- Automate MP Analysis
- iVIREONS webpages (Mike A.)