Mimivirus Relatives in the Sargasso Sea
Elodie Ghedin, Jean-Michel Claverie

To cite this version:
Elodie Ghedin, Jean-Michel Claverie. Mimivirus Relatives in the Sargasso Sea. 2005. hal-00004662

HAL Id: hal-00004662
https://hal.archives-ouvertes.fr/hal-00004662
Preprint submitted on 11 Apr 2005

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Mimivirus Relatives in the Sargasso Sea

Elodie Ghedin\(^1\)
Jean-Michel Claverie\(^2\)

\(^1\) Department of Parasite and Virus Genomics, The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA; Department of Microbiology and Tropical Medicine, George Washington University, Washington DC

\(^2\) Structural and Genomics Information laboratory, CNRS-UPR2589, IBSM, 13402; University of Mediterranee School of Medicine, 13385, Marseille, France

Correspondence to:
Jean-Michel.Claverie@igs.cnrs-mrs.fr
Summary

The discovery and genome analysis of *Acanthamoeba polyphaga Mimivirus*, the largest known DNA virus, challenged much of the accepted dogma regarding viruses. Its particle size (>400 nm), genome length (1.2 million bp) and huge gene repertoire (911 protein coding genes) all contribute to blur the established boundaries between viruses and the smallest parasitic cellular organisms. Phylogenetic analyses also suggested that the Mimivirus lineage could have emerged prior to the individualization of cellular organisms from the three established domains, triggering a debate that can only be resolved by generating and analyzing more data. The next step is then to seek some evidence that Mimivirus is not the only representative of its kind and determine where to look for new Mimiviridae. An exhaustive similarity search of all Mimivirus predicted proteins against all publicly available sequences identified many of their closest homologues among the Sargasso Sea environmental sequences. Subsequent phylogenetic analyses suggested that unknown large viruses evolutionarily closer to Mimivirus than to any presently characterized species exist in abundance in the Sargasso Sea. Their isolation and genome sequencing could prove invaluable in understanding the origin and diversity of large DNA viruses, and shed some light on the role they eventually played in the emergence of eukaryotes.
Introduction
The discovery and genome sequence analysis of Mimivirus [1,2], the largest of the Nucleo-cytoplasmic Large DNA Viruses (NCLDV), challenged much of the accepted dogma regarding viruses. Its particle size (>400 nm), genome length (1.2 million bp) and extensive gene repertoire (911 protein coding genes) all contribute to blur the established boundaries between viruses and the smallest parasitic cellular organisms such as Mycoplasma or Nanoarchea [2]. In the universal tree of life, the Mimivirus lineage appears to define a new branch, predating the emergence of all established eukaryotic kingdoms [2]. Although this result is compatible with various hypotheses implicating ancestral DNA viruses in the emergence of eukaryotes [3-5], it requires confirmation from additional data. An urgent task is thus to convince ourselves that Mimivirus is not the sole representative of its kind (i.e. a viral counterpart to the platypus) and to provide some rational guidance as to where to begin the search for eventual new Mimiviridae.

Mimivirus was serendipitously discovered within Acanthamoeba polyphaga, a free-living ubiquitous amoeba, prevalent in aquatic environments. Phylogenetic analysis of the most conserved genes common to all nucleo-cytoplasmic large double-stranded DNA viruses (NCLDV) [6] positions Mimivirus as an independent lineage, roughly equidistant from the Phycodnaviridae (algal viruses) and Iridoviridae (predominantly fish viruses). Given the ecological affinity of these virus families for the marine environment, we have examined the sequence data set gathered through environmental microbial DNA sampling in the Sargasso Sea [7] to look for possible Mimivirus relatives.

Results
By comparing Mimivirus ORFs to the Sargasso Sea sequence data set and to all other publicly available sequences, 138 (15%) of the 911 Mimivirus ORFs were found to exhibit their closest match (Blastp E-values ranging from $10^{-74}$ to $10^{-4}$ [8]) to environmental sequences (see Materials and Methods). Even before the discovery of Mimivirus, increasingly complex large double-stranded DNA viruses have been isolated, in particular from unicellular algae. The genome analysis of these Phycodnaviruses revealed a variety of genes encoding enzymes from totally unexpected metabolic pathways [9]. Mimivirus added more unexpected genes (such as translation system components [2]) to this list. As the gene repertoire of these large viruses and the gene content of cellular organisms become increasingly comparable, we have to be cautious in the interpretation of environmental/metagenomics sequence data. To focus our study on environmental organisms most likely to be viruses, we limited further analyses to
Mimivirus homologues member of the NCLDV core gene sets [2,6]. These core genes are subdivided into four classes from the most (class I) to least (class IV) evolutionarily conserved [6]. Seven of 10 Mimivirus Class I core genes have their closest homologues in the Sargasso Sea data. This is also the case for 3 of 7 class II core genes, 3 of the 13 class III core genes and 7 of the 16 Class IV core genes (Table 1). Overall, 43% of Mimivirus core genes have their closest homologues in the Sargasso Sea data set. To further assess the viral nature of these unknown microbes, we studied the phylogenetic relationships between the corresponding Mimivirus proteins, their Sargasso Sea homologues, and the closest homologues in other NCLDVs (see Materials and Methods). Figure 1 a-c exhibits three independent phylogenetic trees computed using the MEGA3 software [10] for Mimivirus ORFs R449 (unknown function), R429 (unknown function) and L437 (putative virion packaging ATPase). Figure 1a shows that the closest environmental R449 homologues cluster with Mimivirus separately from the known phycodnaviruses, while other Sargasso Sea homologues cluster in a way suggesting the presence of a new clade distinct from Phycodnaviridae. The tree based on R429 and L437 (Fig. 1b, c) similarly suggests the presence of close Mimivirus relatives not belonging to the Phycodnaviridae or Iridoviridae clades.

Another piece of evidence substantiating the existence of an unknown Mimivirus relative in the Sargasso Sea is the discovery of contigs built from the data that contain multiple genes with a high degree of similarity to Mimivirus genes. A spectacular case is illustrated in Figure 2. Here, a 4.5 kb scaffold (See supplementary information) exhibits 4 putative ORFs. When compared to the whole nr database, each of them has as a best match 4 distinct Mimivirus ORFs: thiol oxidoreductase R368 (29% identical, E-value<10^{-9}), NTPase-like L377 (25% identical, E-value<10^{-20}), unknown function L375 (34% identical, E-value<10^{-30}), and DNA repair enzyme L687 (40% identical, E-value<10^{-62}). Moreover, the gene order is conserved for three of them (R368, L375, L377). Such colinearity is rarely observed between viral genomes except for members of the same family. Unfortunately, the sequences of these genes are not conserved enough to allow the construction of informative phylogenetic trees that would include other NCLDV orthologues.

As of today, genes encoding capsid proteins are among the most unequivocal genes of viral origin. Except for cases of integrated proviral genomes, no cellular homologues of viral capsid proteins have ever been found. During our study, the closest homologues of Mimivirus capsid proteins were found to be capsid protein genes of environmental origin. For example, Mimivirus capsid protein (R441) was found to be 48.5% identical to an unknown environmental sequence, when it is only 36.2% identical to the major capsid protein Vp49 of
Chlorella virus CVG-1, its best match among known viruses (Figure 3). As the environmental capsid protein sequence also shares 44.5% identical residues with the CVG-1 Vp49, the corresponding uncharacterized virus appears to lie at an equal evolutionary distance from the Mimiviridae and the Phycodnaviridae.

Discussion
Our results predict that DNA viruses of 0.1 to 0.8 microns in size exist in the Sargasso Sea that are evolutionarily closer to Mimivirus than to any presently characterized species. These viruses are abundant enough to have been collected by environmental sampling. It is actually expected that many novel viruses will be encountered in natural waters in which they constitute the most abundant microorganisms [11, 12]. There might be as many as 10 billion virus particles per litre of ocean surface waters [13]. Interestingly, the specialized literature abounds of descriptions of large virus-like particle associated with algae [e.g. 14,15,16], or various marine protists [17,18]. With the exception of Phycodnaviruses [19], the genomic characterization of these viruses has not been attempted. Guided by the results presented here, their isolation and genome sequencing could prove invaluable in understanding the diversity of DNA viruses and the role they eventually played in the evolution of eukaryotes.

Materials and Methods
The protocols used to collect Sargasso Sea environmental micro-organisms and generate DNA sequences from these samples has been described elsewhere (7). The data analyzed here correspond to “bacteria-sized” organisms that have passed through 3 µm filters and been retained by 0.8 µm to 0.1 µm filters. Mimivirus-like particles (0.8-0.4 µm) belong in this range.

Database similarity searches were performed using the Blast suite of programs (8) (default options) as implemented on the www.giantvirus.org web server and as implemented at The Institute for Genomic Research. Final similarity searches were performed on the non-redundant peptide sequence databases (nr) and environmental data (env-nr) downloaded from the National Institute for Biotechnology Information ftp server (ftp.ncbi.nlm.nih.gov/blast/db/) on March 14, 2005. To avoid missing potential better matches with annotated virus ORFs, all Mimivirus ORFs exhibiting a best match (blosum62 scoring scheme) in env-nr were also searched against all DNA virus genomes using TblastN (peptide query against translated nucleotide sequence). The comprehensive list of Mimivirus ORFs exhibiting a best match in the env-nr database is given in the supplementary material section.
Phylogenetic analyses were conducted using MEGA version 3.0 (10) (option: Neighbor joining, 250 pseudo-replicates, and gaps handled by pairwise deletion). Tree branches were condensed for bootstrap values < 50%.

Only Mimivirus ORFs with best matching homologues in DNA viruses and belonging to the nucleo-cytoplasmic large DNA virus core gene set (2, 6) were analyzed in detail. These ORFs (and matching status) are listed in Table 1 from the most conserved (type 1, in yellow) to the least conserved (type 2 in green, type 3 in blue, and type 4 in white). Phylogenetic analyses were limited to viral homologues and environmental sequences exhibiting a reciprocal best match relationship with the corresponding Mimivirus ORF (putative orthologues) (YES in the rightmost column). The three cases (red lines in Table 1) exhibiting the best bootstrap values are shown in Figure 1. Cases of complex relationships, for instance due to the presence of many paralogues (e.g. capsid proteins), are also indicated. These cases of non-reciprocal best matches are frequent (i.e. the closest homologue of a Mimivirus ORFs being an environmental sequence, but the latter sequence exhibiting a better match with a different ORF in the nr database).

Two environmental sampling contigs - contig IBEA_CTG_1979672 (AACY01022731, GI:44566181) and contig IBEA_CTG_1979673 (AACY01022732, GI:44566179) - are linked in a 4,465 bp scaffold (scaffold IBEA_SCF=2208413) found to contain four ORFs with strong matches to Mimivirus peptides (R368, L377, L375, and L687). The three colinear ORFs (R368, L377, L375) are found on one contig while the orthologue to Mimivirus ORF L687 is found in the second contig. It is conceivable that the lack of colinearity for this fourth ORF is due to an assembly error.

References
1. La Scola B, Audic S, Robert C, Jungang L, de Lamballerie X, et al. (2003) A giant virus in amoebae. Science 299: 2033.
2. Raoult D, Audic S, Robert C, Abergel C, Renesto P, et al. (2004) The 1.2-megabase genome sequence of Mimivirus. Science 306: 1344-1350.
3. Villarreal LP, DeFilippis VR (2000) A hypothesis for DNA viruses as the origin of eukaryotic replication proteins. J. Virol. 74: 7079-7084.
4. Bell PJ (2001) Viral eukaryogenesis: was the ancestor of the nucleus a complex DNA virus? J. Mol. Evol. 53: 251-256.
5. Takemura M. (2001) Poxviruses and the origin of the eukaryotic nucleus.
6. Iyer L.M. Aravind L., Koonin E.V. (2001) Common origin of four diverse families of large eukaryotic DNA viruses. J. Virol. 75: 11720-11734.

7. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, et al. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. Science 304: 66-74.

8. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402.

9. Van Etten JL (2003) Unusual life style of giant chlorella viruses. Annu Rev Genet 37:153-195.

10. Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform. 5: 150-163.

11. Wommack KE, Colwell RR (2000) Virioplankton: viruses in aquatic ecosytems. Microbiol Mol Biol Rev 64: 69-114.

12. Ghedin E, Fraser C (2005) A virus with big ambitions. Trends Microbiol 13: 56-57.

13. Fuhrman JA. (1999) Marine viruses and their biogeochemical and ecological effects. Nature 399: 541–548.

14. Mattox KR, Stewart KD (1972). Probable virus infections in four genera of green algae. Can. J. Microbiol. 18, 1620-1621.

15. Dodds JA, Cole A (1980) Microscopy and biology of Uronema gigas, a filamentous eukaryotic green alga, and its associated tailed virus-like particle. Virology 100: 156-165.

16. Van Etten JL, Lane LC, Meints RH (1991) Viruses and viruslike particles of eukaryotic algae. Microbiol Rev 55: 586-620.

17. Shin W, Boo SM (1999) Virus-like particles in both nucleus and cytoplasm of Euglena viridis. Algological Studies 95: 125-131.

18. Growing MM (1993) LargeVLPs from vacuoles of phaeodarian radiolarians and from other marine samples. Marine Ecology progress Series 101: 33-43.

19. Van Etten JL, Graves MV, Muller DG, Boland W, Delaroque N (2002) Phycodnaviridae--large DNA algal viruses. Arch Virol 147:1479-516.

Acknowledgements
We are indebted to James van Etten for pointing out some ancient observations of very large virus-like particles in algae and marine protists. We thank Stéphane Audic for his help with the www.giantvirus.org server and Hiroyuki Ogata and Vish Nene for reading the manuscript.
| ORF# | Definition                                                                 | Best score in nr | Best score in DNA viruses | Best score in Sargasso Sea | Status      | Reciprocal Best match |
|------|---------------------------------------------------------------------------|------------------|---------------------------|---------------------------|-------------|------------------------|
| L206 | Helicase III / VV D5                                                    | 167- virus       | 167                       | 214                       | Best ENV    | YES                    |
| R322 | DNA pol (B family) extein                                               | 207              | 167                       | 238                       | Best ENV    | YES                    |
| L437 | A32 virion packaging ATPase                                             | 169 - virus      | 169                       | 191                       | Best ENV    | YES                    |
| L396 | VV A18 helicase                                                         | 200 - virus       | 200                       | 187                       | -           |                        |
| L425 | Capsid protein                                                          | 119 - virus       | 119                       | 142                       | Best ENV    | complex                |
| R439 | Capsid protein                                                          | 164 - virus       | 159                       | 173                       | Best ENV    | complex                |
| R441 | Capsid protein                                                          | 137 - virus       | 147                       | 209                       | Best ENV    | complex                |
| R596 | E10R-Thiol oxidoreductase                                               | 104 -virus        | 105                       | 119                       | Best ENV    | complex                |
| R350 | F10L - prot. Kinase                                                     | 86 - virus        | 86                        | 58                        | -           |                        |
| R393 | A1L-transcr. factor                                                     | 52 - virus        | 47                        | 65                        | Best ENV    |                        |
| L324 | Mu1-like NTP PP-hydrolase                                               | 40               | 38                        | 39                        | -           |                        |
| L323 | Myristoylated virion prot. A                                            | 92               | 42                        | 154                       | Best ENV    |                        |
| R439 | PCNA                                                                     | 76               | 74                        | 740                       | -           |                        |
| L312 | Small Ribonucl. reduct                                                  | 341              | 338                       | 310                       | -           |                        |
| R313 | Large Ribonucl. reduct                                                  | 766              | 741                       | 740                       | -           |                        |
| R429 | PBCV1-A494R-like                                                       | 152-virus         | 152                       | 216                       | Best ENV    | YES                    |
| L37  | BroA, KilA-N                                                            | 123-virus         | 124                       | 65                        | -           |                        |
| R382 | mRNA–capping enz.                                                       | 86               | 78                        | 166                       | Best ENV    | YES                    |
| L244 | RNA pol. sub 2 (Rbp2)                                                   | 727              | 416                       | 508                       | -           |                        |
| R501 | RNA pol. sub.1 (Rbp1)                                                   | 805              | 415                       | 520                       | -           |                        |
| R195 | ESV128- Glutaredoxin                                                    | 50               | 39                        | 49                        | -           |                        |
| R622 | S/Y phosphatase                                                         | 75               | 73                        | 65                        | -           |                        |
| R311 | ClIV193R BIR domain                                                    | 68               | 44                        | 51                        | -           |                        |
| L65  | Virion memb. prot.                                                      | 44               | 44                        | -                         | -           |                        |
| R480 | Topoisomerase II                                                        | 902              | 717                       | 367                       | -           |                        |
| L221 | Topoisomerase I bacterial                                               | 528              | 35                        | 516                       | -           |                        |
| R194 | Topoisomerase I pox-like                                               | 188              | 100                       | 145                       | -           |                        |
| L364 | SW1/SNF2 helicase                                                       | 70-virus         | 70                        | 72                        | Best ENV    | YES                    |
| L4 + 7 par | N1R/P28 DNA binding prot                                        | 123-virus       | 124                       | 72                        | -           |                        |
| L540 | Pre-mRNA helicase - splicing                                           | 256              | 136                       | 214                       | -           |                        |
| L235 | RNA pol subunit5                                                        | 69               | 38                        | 50                        | -           |                        |
| R354 | Lambda-type exonuclease                                                 | 69-virus         | 69                        | 154                       | Best ENV    | YES                    |
| R343 | RNAase III                                                              | 129              | 112                       | 131                       | Best ENV    | YES                    |
| R141 | GDP mammose 4.6-dehydratase                                             | 294              | 68                        | 252                       | -           |                        |
| L258 | Thymidine kinase                                                        | 151              | 140                       | 124                       | -           |                        |
| L271 + par | Ankyrin repeats (66 paralogs)                                      | 179              | 152                       | 192                       | Best ENV    | complex                |
| R325 | Metal-dependent hydrolase                                               | 69-virus         | 69                        | 105                       | Best ENV    | YES                    |
| L477 | Cathepsin B                                                             | 226              | 43                        | 47                        | -           |                        |
| R497 | Thymidylate synthase                                                    | 278              | 242                       | 217                       | -           |                        |
| R449 | Uncharacterized prot.                                                   | 69-virus         | 69                        | 129                       | Best ENV    | YES                    |
| R303 | NAD-dependent DNA ligase                                                | 270-virus        | 270                       | 228                       | -           |                        |
| L805 | MACRO domain                                                            | 36               | 33                        | -                         | -           |                        |
| R571 | Patatin-like phospholipase                                              | 105              | 80                        | 122                       | Best ENV    | YES                    |
| L446 | Uncharacterized prot.                                                   | 48-virus         | 48                        | 65                        | Best ENV    | YES                    |
Figure legend

**Figure 1.** Phylogenetic evidence of uncharacterized Mimivirus relatives. (a) Neighbor-joining (NJ) clustering (see Materials and Methods) of Mimivirus R449 ORF with its best matching (~35% identical residues) environmental homologues (noted Sargasso1 to Sargasso6 according to their decreasing similarity) and closest viral orthologues (28% identical). (b) NJ clustering of Mimivirus R429 ORF with its best matching (~50% identical) environmental homologues (noted Sargasso1 to Sargasso5) and closest viral orthologues (35% identical). (c) NJ clustering of Mimivirus putative virion packaging ATPase L437 with its best matching (~45% identity) environmental homologues (Sargasso1 and Sargasso2) and closest viral orthologues (34% identical). Abbreviations: Phyco: Phycodnavirus; PBCV: *Paramecium bursaria* chlorella virus 1; EsV: *Ectocarpus siliculosus* virus; FsV: Feldmannia sp. virus; HaV: Heterosigma akashiwo virus; Irido: Iridovirus; LCDV: Lymphocystis disease virus 1; Frog: Frog virus 3; Amby: *Ambystoma tigrinum stebbensi* virus; Rana: *Rana tigrina ranavirus*; Chilo: Chilo iridescent virus. Bootstrap values larger than 50% are shown. Branches with lower values were condensed.

**Figure 2.** Organization of four Mimivirus ORF best matching homologues in a 4.5 kb environmental sequence scaffold (approximately to scale). The three colinear Mimivirus homologues are in green. Unmatched ORF extremities are indicated by dots. The two diagonal lines indicate where the two contigs are joined on the scaffold.

**Figure 3.** Partial 3-way alignment (N-terminus region) of Mimivirus capsid protein (R441) with its best matching homologues in the NR and Environmental sequence databases. The Mimivirus R441 protein shares 83/229 (36.2%) identical residues (colored in red or blue) with the major capsid protein Vp49 of Chlorella virus CVG-1 and 111/229 (48.5%) identical residues (indicated in red or green) with the N-terminus of a capsid protein from an unknown large virus sampled from the Sargasso Sea (Accession: EAD00518). On the other hand, the CVG-1 Vp49 and the Sargasso Sea sequence share 44.5 % identical residues. By comparison, the CVG-1 Vp49 protein share 72% of identical residue with PBCV-1 Vp54, its best matching homologue among known phycodnaviruses.
Figure 1

a

b

Sargasso1
Sargasso5
Mimi R429
Sargasso4
Sargasso6
Sargasso3
Sargasso2
Sargasso1
Mimi R449

Irido-Frog
Irido-Rana
Irido-Amby
Irido-Chilo
Mimi L437
Sargasso1
Sargasso2
Phyco-EsV
Phyco-FsV
Phyco-IIaV
Phyco-PBCV

Sargasso1
Sargasso5
Mimi R429
Sargasso3
Sargasso4
Sargasso2
Phyco-PBCV
Phyco-EsV
Phyco-FsV
Irido-Chilo
Irido-LCDV

Phyco-PBCV
Phyco-EsV
Phyco-FsV
Phyco-HaV

Phyco-PBCV
Phyco-EsV
Phyco-FsV
Phyco-HaV
Figure 2
Figure 3

CVG1-vp49       MAGGLSQLVAYGAQDVYLTPQITFFKTVYRRYTNFAVESIQTINGSV
MIMI-R441       MAGGLSQLVAYGAQDVYLTPQITFFKTVYRRYTNFAVESIQTINGSV
Sargasso1       MGGLQLVAYGAQDVYLTPQITFFKTVYRRYTNFAVESIKQTNGTA

CVG1-vp49       GFGNKVSTQISRNGDLITDIVVEFVLTQGPTFY----------------
MIMI-R441       DFGSTVSCSTLSKSDMIKINEIVYIELPSVNFYDESG------NLDKFKK
Sargasso1       DFGKKSCTISRNGDLVHRIFLQTTPAQKYDYASAGGTVYNSNSMK

CVG1-vp49       ---------CAEQLLQDVTEISGQRIDKYADWFRMYDSLFRMD-----
MIMI-R441       ---FATVRNIGAIKAVSIEEGKLIDKQYGEWMYWQNTNKS--DEG
Sargasso1       DGILRWINWVGEKLYAEIEIQQRIDKLYGEWHLHWHQLTNTASHDEG

CVG1-vp49       --NDRQNYRRM-----TDFVNDEPAVTAKRFYYVPLIFFNQTPGLALPLI
MIMI-R441       LDKMIKNIPL-----YDFSNKCP-----KSFLYVPLEFWFCRNSGLSLPLV
Sargasso1       YQRVMGNPALTNNSTNTVAGAEIAQDLYVPLQFWFCRNPGLALPLI

CVG1-vp49       ALQYHSVKEVLYFTLAST-----------------VNGITAVEGGAATAVAP
MIMI-R441       ALSSSEVKITIFRSADECYRIGPDSIEIMEDIVPSFEFGDYIEKIG
Sargasso1       ALQYHVEKINIEFEEL-----------------KNLFIADKTTAATAVTN
Supplementary material

List of Mimivirus ORFs exhibiting a best match in the env-nr database
| Mimi ORF | Best Env-nr Match ID | Score | E-value | Best nr Match ID | Score | E-value |
|----------|----------------------|-------|---------|------------------|-------|---------|
| L18      | gi|43710841|gb|EAF10508.1| 137.1 | 4.7e-31 | gi|51244078|ref|YP_063962.1| 106.7 | 2.7e-21|
| L25      | gi|42923128|gb|EAB28610.1| 169.1 | 5e-41 | gi|34876677|ref|XP_214012.2| 163.3 | 1.1e-38|
| L75      | gi|43124292|gb|EAC28017.1| 51.6 | 1.1e-05 | gi|13794513|gb|AAK39888.1| 51.22 | 5.5e-05|
| L93      | gi|42923128|gb|EAB28610.1| 177.9 | 1.6e-43 | gi|40740451|gb|EAA59641.1| 156.8 | 1.5e-36|
| L100     | gi|42923128|gb|EAB28610.1| 119 | 1.4e-25 | gi|40740451|gb|EAA59641.1| 117.9 | 1.2e-24|
| L102     | gi|44215848|gb|EAH75003.1| 46.98 | 0.00017 | gi|19703853|ref|NP_603415.1| 41.2 | 0.353|
| R106     | gi|43776260|gb|EAF42614.1| 139 | 8.1e-32 | gi|37676542|ref|NP_936938.1| 107.1 | 1.4e-21|
| L111     | gi|43296731|gb|EAD14168.1| 46.21 | 0.0006 | gi|16805307|ref|NP_473335.1| 41.2 | 0.777|
| L113     | gi|43770287|gb|EAF39597.1| 50.45 | 8.8e-06 | gi|23498951|emb|CAD51029.1| 43.9 | 0.003|
| R118     | gi|44281621|gb|EAI17665.1| 187.6 | 1.6e-46 | gi|56965351|ref|NP_936938.1| 107.1 | 1.4e-21|
| R132     | gi|42849963|gb|EAA92307.1| 98.6 | 4.3e-20 | gi|33864966|ref|NP_896525.1| 61.62 | 2.1e-08|
| R135     | gi|44399015|gb|EAI99687.1| 99.37 | 1.4e-19 | gi|48769830|ref|ZP_00274174.1| 97.06 | 2.7e-18|
| L136     | gi|43770287|gb|EAF39597.1| 50.45 | 8.8e-06 | gi|13794513|gb|AAK39888.1| 51.22 | 5.5e-05|
| R143     | gi|47888785|gb|EAF49393.1| 101.7 | 8e-21 | gi|54031509|ref|NP_00363643.1| 53.91 | 7.4e-06|
| L174     | gi|43770436|gb|EAF09806.1| 48.52 | 0.0001 | gi|13358136|ref|NP_078410.1| 45.44 | 0.003|
| L177     | gi|43660484|gb|EAE85189.1| 47.75 | 0.00021 | gi|56526423|emb|CAD51029.1| 43.9 | 0.003|
| L193     | gi|43116621|gb|EAC24307.1| 138.7 | 1.7e-31 | gi|56965351|ref|NP_896525.1| 61.62 | 2.1e-08|
| L206     | gi|44511459|gb|EAC77190.1| 210.7 | 2e-53 | gi|33864966|ref|NP_896525.1| 61.62 | 2.1e-08|
| L207     | gi|43582640|gb|EAE46140.1| 115.2 | 1.9e-24 | gi|13358136|ref|NP_078410.1| 45.44 | 0.003|
| L208     | gi|43154232|gb|EAC42788.1| 54.3 | 6.8e-07 | gi|12350912|ref|NP_701792.1| 44.67 | 0.002|
| L215     | gi|43198765|gb|EAC64922.1| 45.44 | 9.8e-05 | gi|30249486|ref|NP_841556.1| 43.9 | 0.001|
| R240     | gi|44090491|gb|EAG93597.1| 83.57 | 9.3e-15 | gi|30409752|gb|AAP32727.1| 55.45 | 2e-06|
| L250     | gi|43527692|gb|EAE18541.1| 133.7 | 3.8e-30 | gi|17313675|ref|NP_476888.1| 121.3 | 7.9e-26|
| L251     | gi|44049491|gb|EAB84927.1| 263.5 | 8.5e-69 | gi|13358136|ref|NP_078410.1| 45.44 | 0.003|
| R267     | gi|42923128|gb|EAB28610.1| 114 | 2.9e-24 | gi|40740517|gb|EAA59707.1| 111.7 | 5.7e-23|
| L279     | gi|42977843|gb|EAB55869.1| 46.21 | 0.00082 | gi|23508123|ref|NP_700801.1| 45.05 | 0.007|
| R296     | gi|44416070|gb|EAE11899.1| 120.7 | 6.8e-32 | gi|13358136|ref|NP_078410.1| 45.44 | 0.003|
| R299     | gi|43651599|gb|EAE80654.1| 82.42 | 3.4e-15 | gi|46440930|gb|EAL00231.1| 80.88 | 3.7e-14|
| L300     | gi|44143264|gb|EAE36242.1| 59.31 | 1.5e-08 | gi|150292185|ref|XP_448525.1| 54.3 | 1.7e-06|
| R301     | gi|42942338|gb|EAB38176.1| 64.7 | 1.8e-09 | gi|15042460|gb|AAP32727.1| 48.14 | 0.00068|
| L315     | gi|42953235|gb|EAB43598.1| 88.97 | 5.4e-17 | gi|56964473|ref|YP_176204.1| 70.86 | 5.9e-11|
|gi|44068613|gb|EAG95539.1| 150.6 1.2e-35 gi|24378003|gb|AAN59275.1| 144.1 4.4e-33 |
|gi|44283592|gb|EAI18988.1| 152.5 6e-36 gi|37519905|ref|NP 923282.1| 104.4 7.4e-21 |
|gi|43163235|gb|EAC47212.1| 119.4 8.9e-27 gi|49235486|ref|ZP 00329554.1| 115.5 3.2e-25 |
|gi|44016225|gb|EAG66729.1| 69.32 2.4e-11 gi|15921096|ref|NP 376765.1| 66.24 7.4e-10 |
|gi|44283592|gb|EAI18988.1| 152.5 6e-36 gi|37519905|ref|NP 923282.1| 104.4 7.4e-21 |
|gi|43163235|gb|EAC47212.1| 119.4 8.9e-27 gi|49235486|ref|ZP 00329554.1| 115.5 3.2e-25 |
|gi|44016225|gb|EAG66729.1| 69.32 2.4e-11 gi|15921096|ref|NP 376765.1| 66.24 7.4e-10 |