Mass Spec data available on most of the websites, not yet available for MicrosporidiaDB or PiroplasmaDB

Mass Spectrometry at EuPathDB

- AmoebaDB
- CryptoDB
- GiardiaDB
- MicrosporidiaDB
- PiroplasmaDB
- PlasmoDB
- ToxoDB
- TrichDB
- TriTrypDB

EuPathDB Workshop
June 2011
Mark Heiges <mheiges@uga.edu>

MS Use Cases

- evidence for protein expression

- evidence for subcellular localization

- confirm gene structural annotation

- detect missed gene annotations

Tandem Mass Spectrometry (MS/MS)

2D Gel

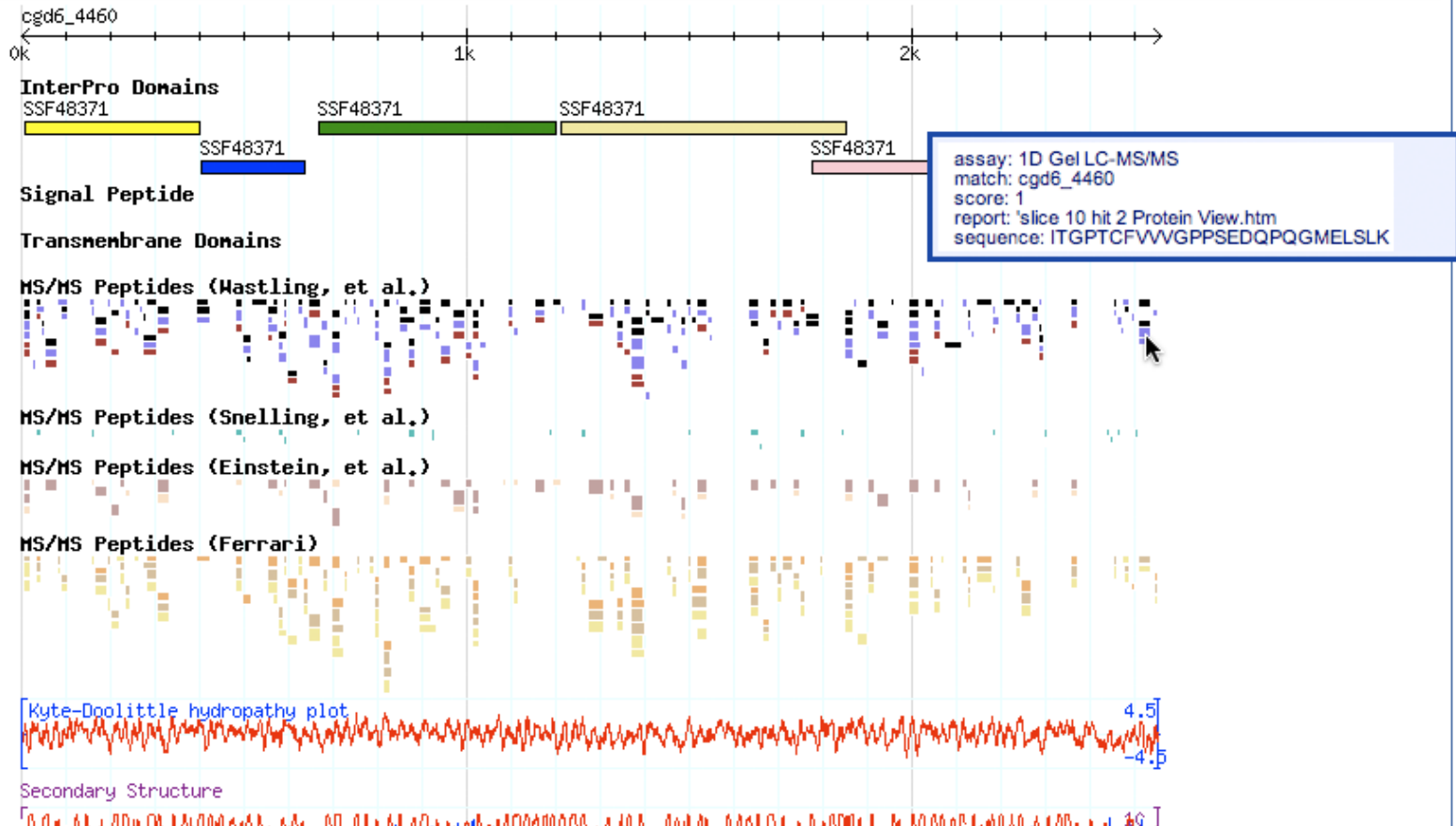Proteolysis
(trypsin)

Separate by HPLC
(solubility)
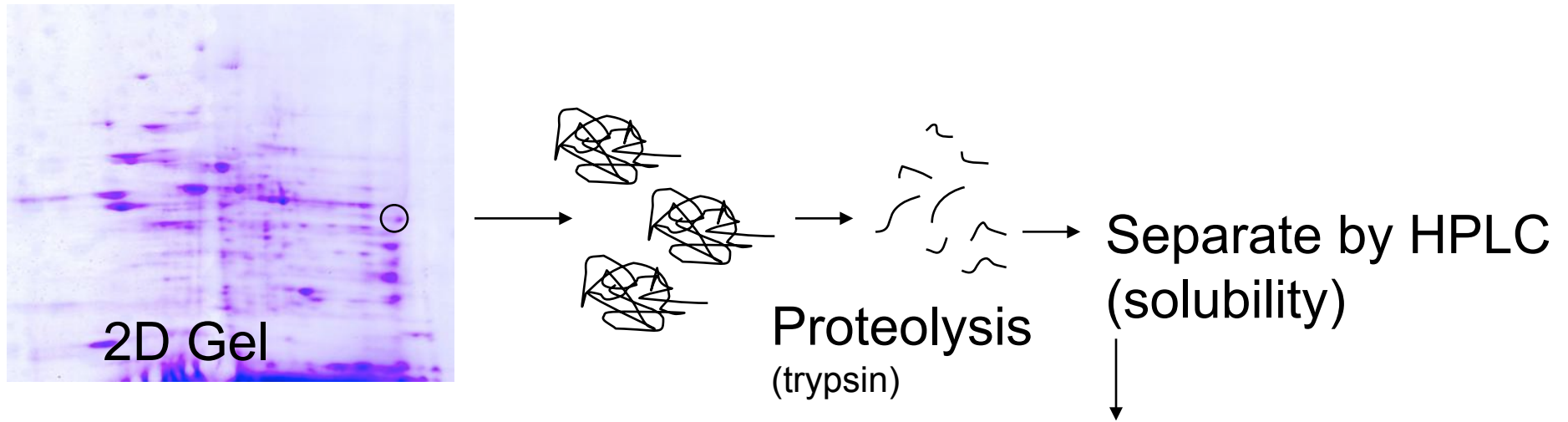
One example MS methodology for protein analysis is presented. There are many other techniques that vary in sample preparation and instruments used.
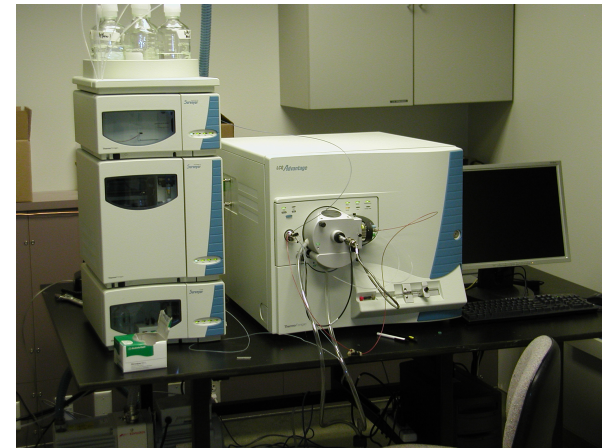
1. Separate by peptide mass (MS)
2. Fragment, analyze component masses (MS)

Predicted Protein Features [Data Sources]

assay: 1D Gel LC-MS/MS
match: cgd6_4460
score: 1
report: 'slice 10 hit 2 Protein View.htm
sequence: ITGPTCFVVVGPPSEDQPQGMELSLK

example of peptides discovered through MS analysis, represented graphically in GBrowse on our site.

The glyphs are color coded by assay type. Pointing at one of the glyphs will trigger a popup with information about the specific peptide

This protein was identified by three different analyses (blue, red, black colors) by Wastling and three different analyses by Snelling (aqua, peach, red colored glyphs) and so on for the other data providers. This is repeated discovery of the same protein by so many strategies and independent researcher provides very good indication that the protein was correctly identified by the MS searches.

ORFs with
Mass Spectrometry Evidence

- CryptoDB
- GiardiaDB
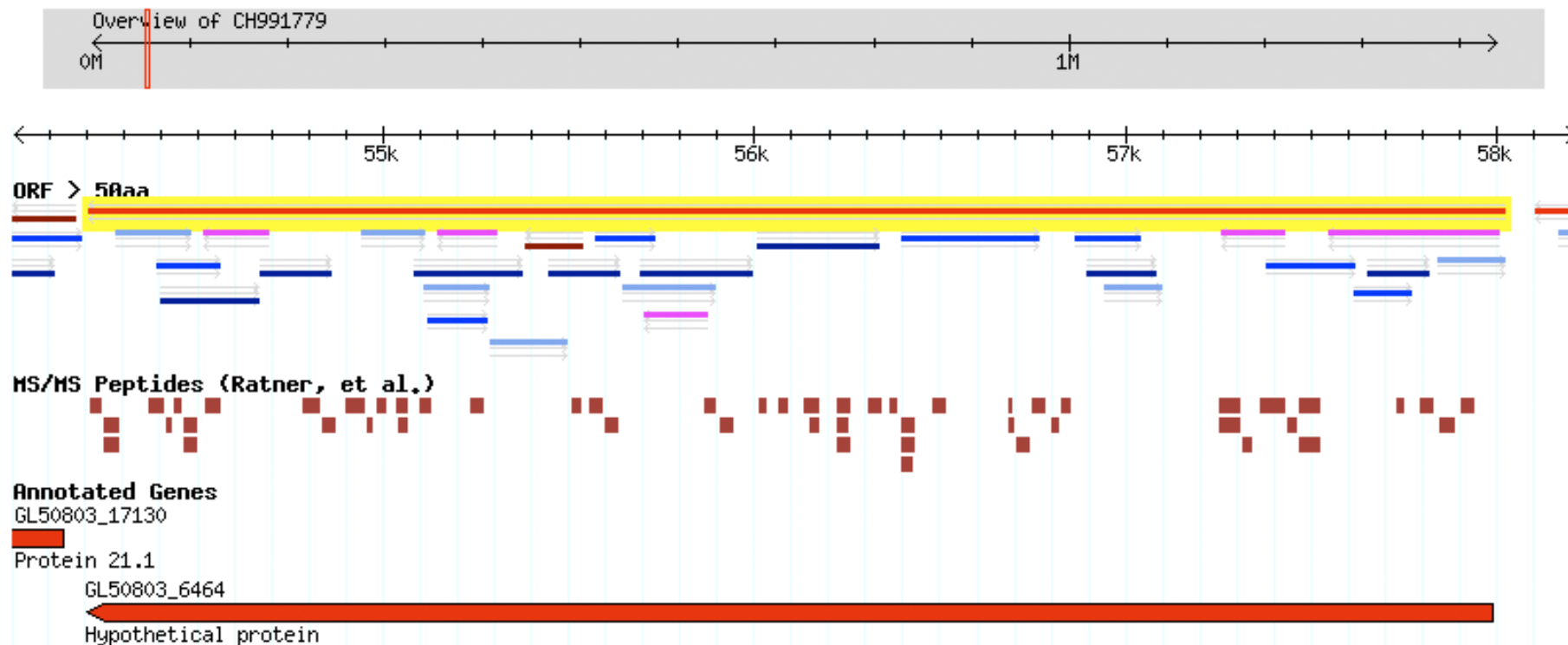- ToxoDB

ORFs at EuPathDB

AC**TGA**TCCAGCTCAGTACATTTTAGATACTTCTATTTATTCATCTGG**TGA**TTTA

S   S   S   V   H   F   R   Y   F   Y   L   F   I   W

- computed in all 6 reading frames
- do not necessarily begin with Met
- are not annotated or reviewed

Re-Mapping Peptide ORF Hits to Genes

MS analysis often includes peptides identified by spectra searches against an ORF database. Frequently those ORFs correspond to annotated genes, as shown here. Such ORFs are typically not very interesting to our community. So, in these cases we move the MS identified peptides from the ORF to the annotated protein. In the example shown here, the peptides shown were orginally assigned, by the data provider, to the ORF highlighted in yellow. We moved the peptides to the GL50583_6464 gene. A search for 'genes with mass spec evidence' will return the annotated gene. A search for 'orfs with mass spec evidence' will NOT return this ORF because we have taken the MS evidence away from the ORF and put it on the gene.
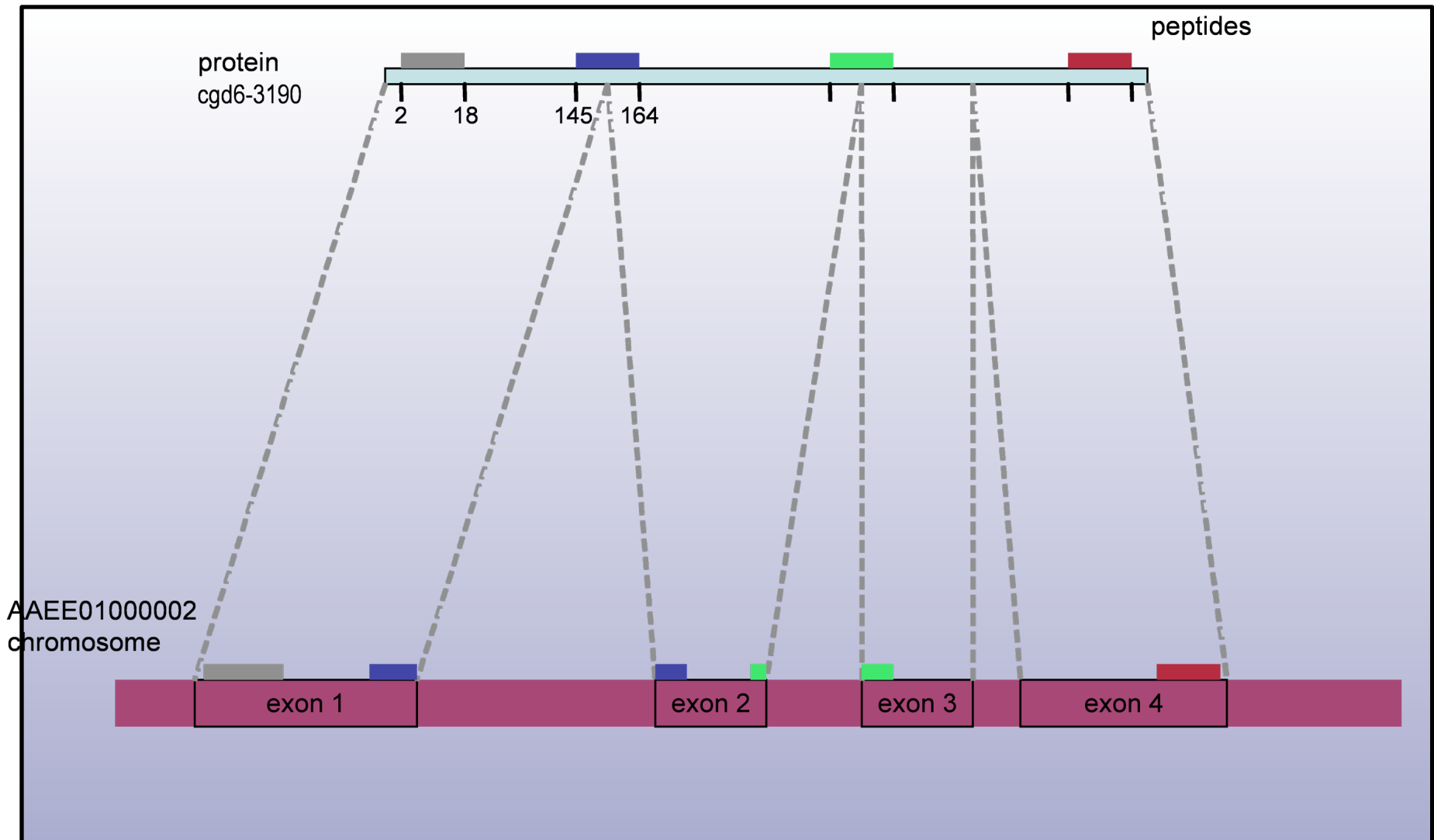
Re-Mapping ORF Peptide Hits to Genes or Not
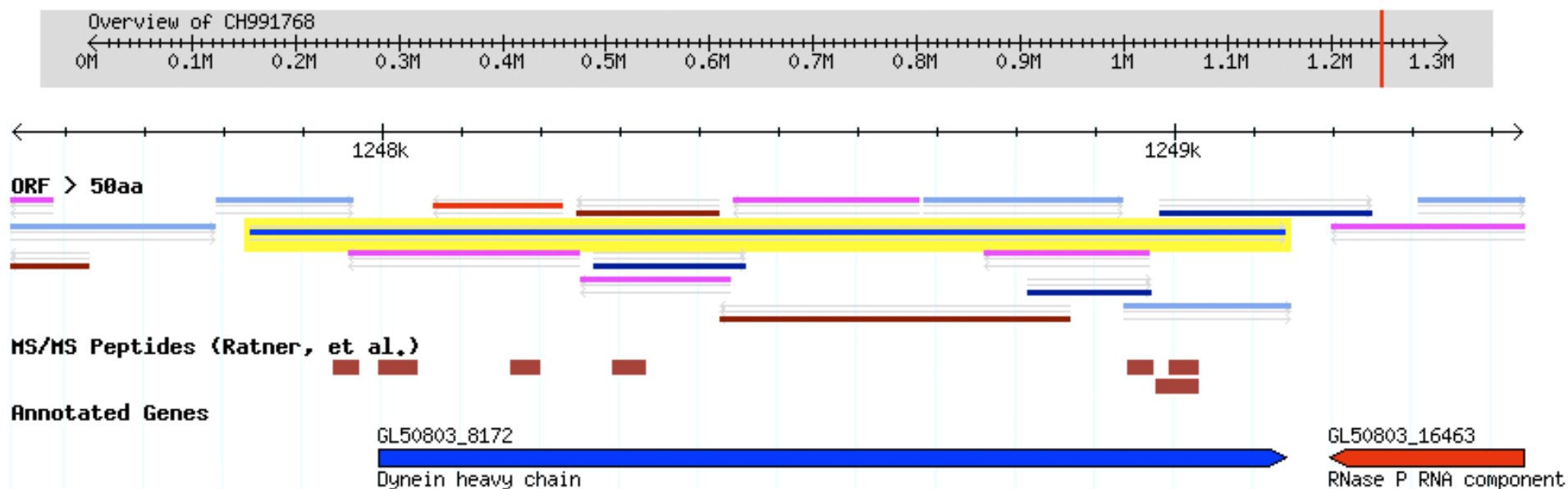
In the example shown here, the peptides shown were orginally assigned, by the data provider, to the ORF highlighted in yellow. Because one of the peptides identified on the ORF does not correspond the the GL5083_8172 gene, we leave all the assigned peptides associated with the orginal ORF. A search for ORFs with MS evidence WILL return this ORF. By moving the peptides from ORFs that are fully contained within an annotated gene (previous slide), the search for possibly interesting identifications, like this example, is made easier. However, be very cautious about attempting to reannote a gene on the sole basis of a single MS peptide, as shown here. That one peptide may have been mis-identified by the MS software.

The data we get from the research labs includes the peptide coodinates on the protein sequence. We are able to calculate the corresponding genomic coordinates for the DNA sequence that encodes the peptide. In some cases those sequences span an intron. Because the gene exons have to be spliced to generate the detected peptide sequences, such examples support the annotated intron/exon boundaries.

Map Peptides to Genome

peptides

protein
cgd6-3190

2   18   145   164

AAEE01000002
chromosome

exon 1   exon 2   exon 3   exon 4

Support for Intron Boundaries
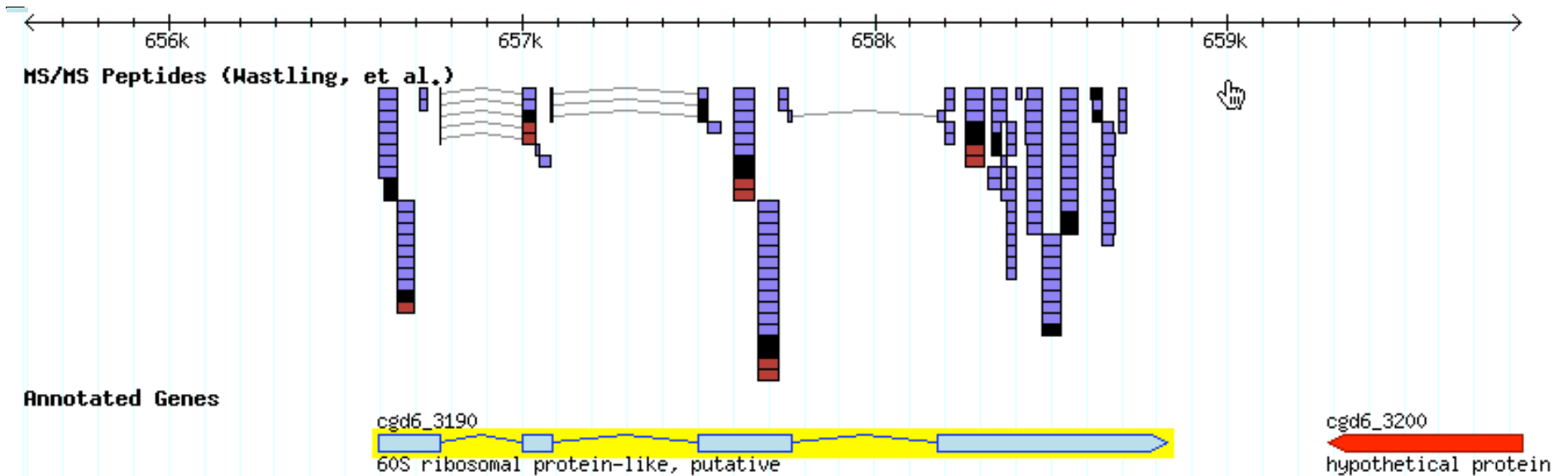
some examples where the DNA sequences encoding an MS peptide span an intron. This support the intron/exon boundaries annotated correctly.

Search using a sequence name, gene name, locus, or other landmark. The wildcard character * is allowed. To center on a location, click the ruler. Use the Scroll/Zoom buttons to change magnification and position.
Examples: AAEE01000001:1-35000, AAEE01000002:106357-116356, AAEL01000026:7652-27651, cgd6_780, CMU_011540.

[Hide banner] [Bookmark this] [Link to Image] [High-res Image] [Help] [Reset]

Search

Landmark or Region:
CM000434:1051852..1061851   (Search)

Reports & Analysis:
[Annotate Restriction Sites ▲▼]   (Configure...)   (Go)

Data Source
[CryptoDB ▲▼]

Scroll/Zoom: « ‹ ⊖ [Show 10 kbp ▲▼] ⊕ › » ☐ Flip

Select the Data Source of your choice (above).

Overview

Overview of CM000434

0M   0.1M   0.2M   0.3M   0.4M   0.5M   0.6M   0.7M   0.8M   0.9M   1M   1.1M   1.2M   1.3M

Details

1052k   1053k   1054k   1055k   1056k   1057k   1058k   1059k   1060k   1061k

MS/MS Peptides (Wastling, et al.)

MS/MS Peptides (Snelling, et al.)

Mass Spec

Experiment:	assay: _16May2006_1D_gel
Sequence:	ILAEDPTGAQYVAQTPMGDISTIISTIK
Description:	match: cgd6_4460
	score: 97.44 (expectation: 1.2e-07)
Number of Matches:	42

MS/MS Peptides (Einstein)

MS/MS Peptides (Ferrari)

Annotated Genes
cgd6_4440   cgd6_4460

cgd6_4450

Clear highlighting

Genome Browser view. This is similar to what was seen on the gene record page but here you have more control over which tracks and specific genome sequence range are shown. To keep the display reasonably compact, we do not display redundant peptide sequences within a given experiment. Instead we show a single block representing the peptide and in a mouseover popup we report the "Number of Matches" seen for that sequence.

Update Image

Reveal Missing Gene Annotations
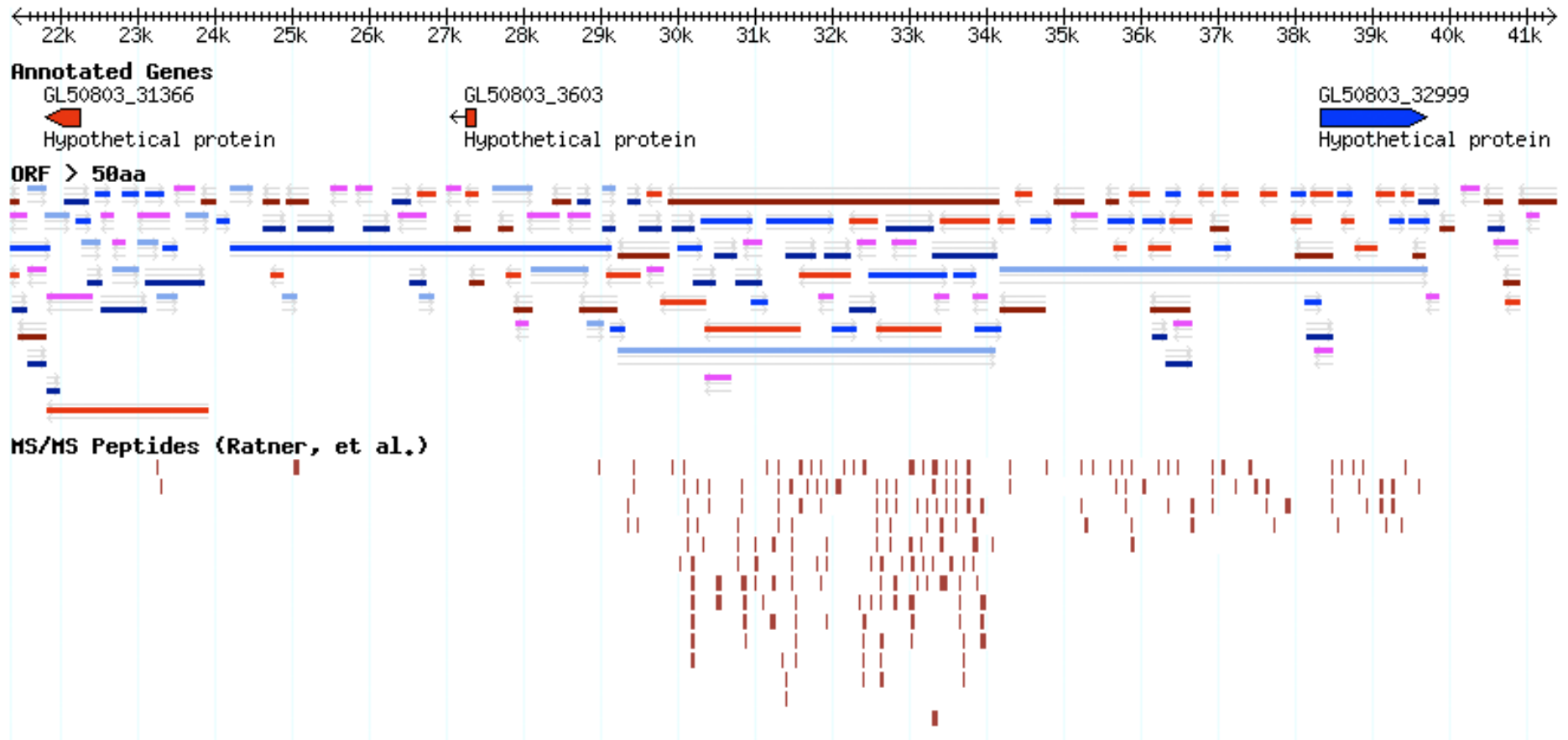
This is a region of a Giardia chromosome where long open reading exist but there is no gene annotated or the annotated gene does not extend for the length of the ORF. There were many peptides identified for this region by searching spectra against the ORF sequences suggesting that the associated ORFs are part of a gene that has not yet been annotated.