



Interactive and computational access *to proteomics data in the PRIDE database*



Daniel RIOS
PRIDE software developer

PRIDE team, Proteomics Services Group
PANDA group
European Bioinformatics Institute
Hinxton, Cambridge
United Kingdom



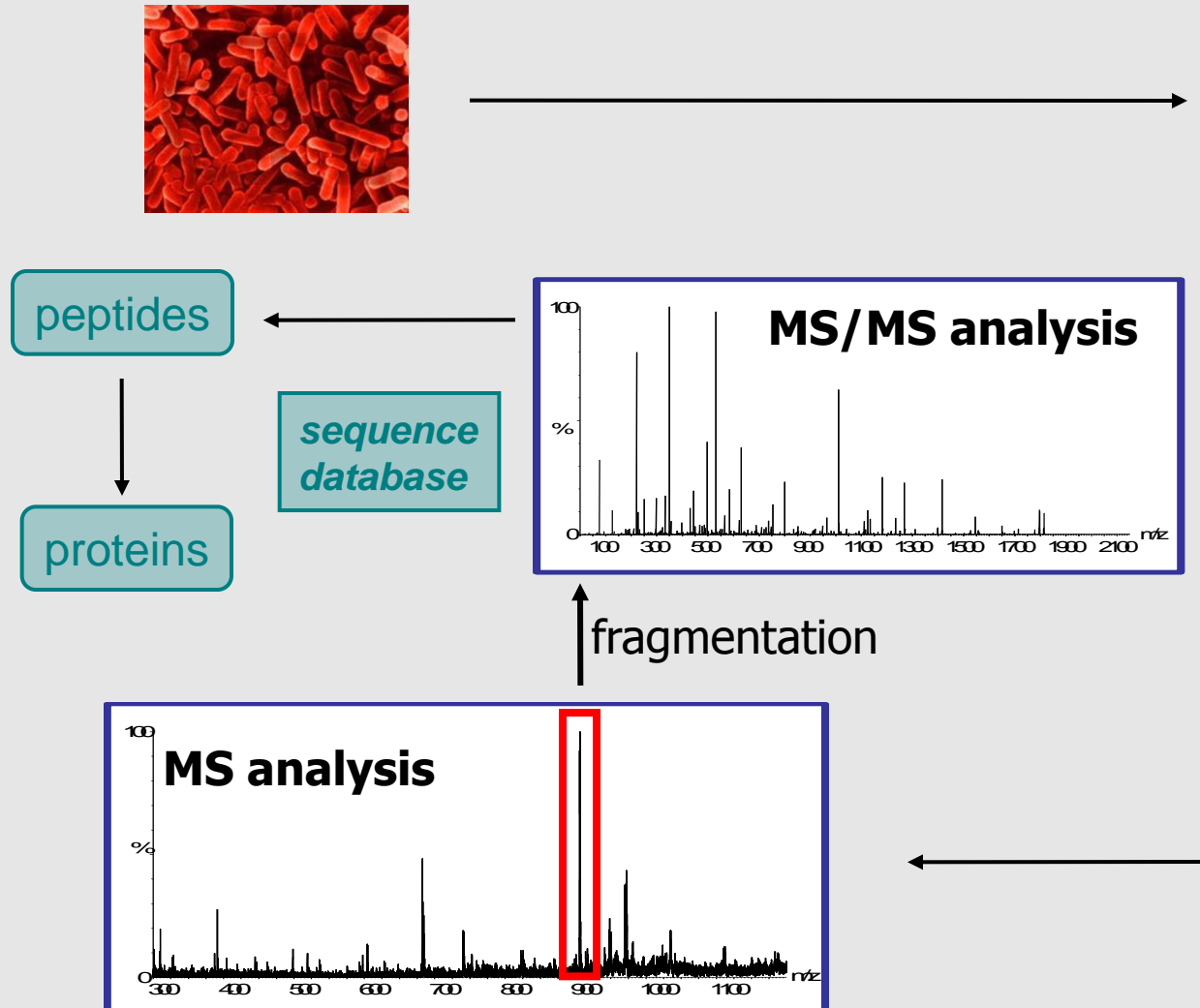
Overview ...

- PRIDE: what do we store ?
- Accessing the data:
 - BioMart
 - DAS
 - DASTY client

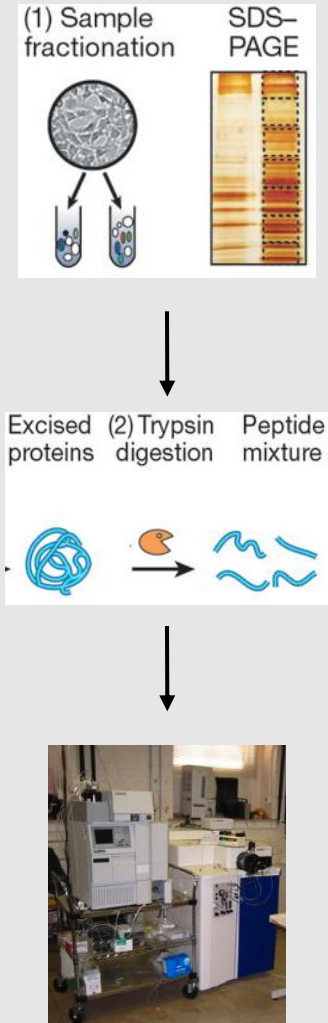
PRIDE



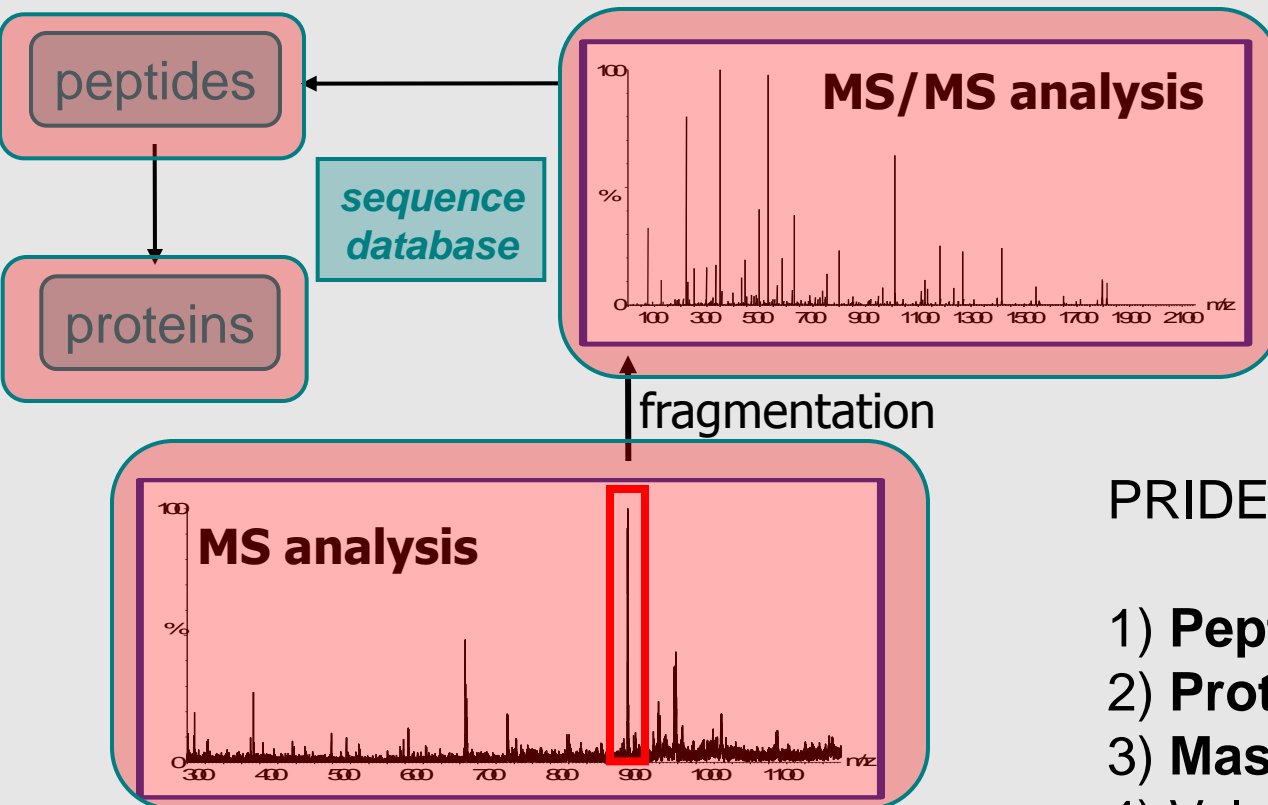
MS proteomics: overall workflow



P
R
O
T
O
C
O
L



PRIDE database (www.ebi.ac.uk/pride)



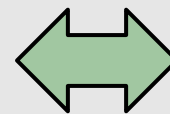
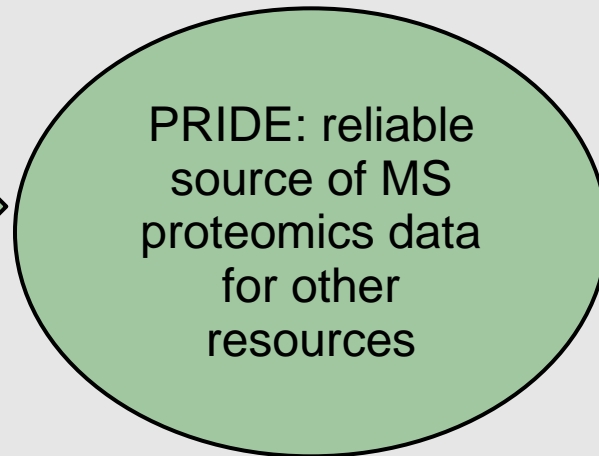
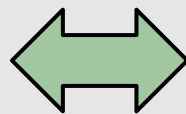
PRIDE stores:

- 1) Peptide IDs
- 2) Protein IDs
- 3) Mass spectra as peak lists
- 4) Valuable additional **metadata**

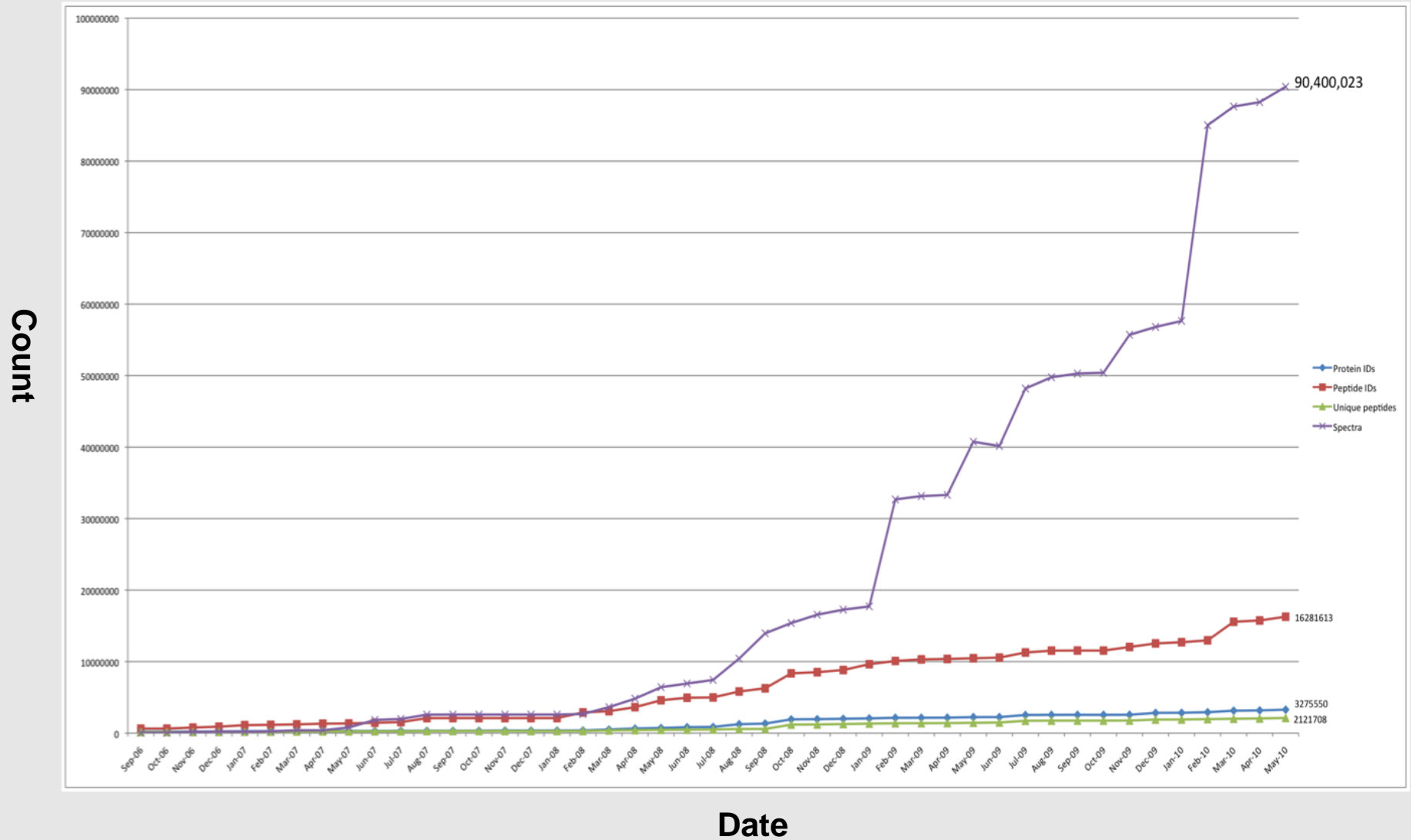
PRIDE: why is it there?



- Repository to support publications (proteomics MS derived data)
- Source of proteomics data for other data resources



PRIDE growth



BIOMART Interface

(<http://www.ebi.ac.uk/pride/prideMart.do>)

BioMart

● A collaboration

- European Bioinformatics Institute (EBI)
- Cold Spring Harbor Laboratory (CSHL)
- Ontario Institute for Cancer Research (OICR)


● Aim



- To develop a structured data query engine that works for biological research

● All benefits of structured data in the relational database but with



- Queries without the knowledge of the table structure
- Scales for big datasets
- Solves integration

PRIDE BioMart (<http://www.ebi.ac.uk/pride/prideMart.do>)

EMBL-EBI  All Databases

Databases Tools EBI Groups Training Industry About Us Help  

EBI > Databases > Proteomics > PRIDE > PRIDE BioMart

Dataset
PRIDE

Filters
Filter by Project : 2D reference map of *Herbaspirillum seropedicae*

Attributes
[None selected]

Dataset
[None Selected]

Please restrict your query using criteria below

Filter by Experiment

Filter by Project ID


PP-1
PP-10
PP-11
PP-12
PP-13



Filter by Project

Q03d27
A simple strategy for straightforward proteome analysis by a combination of Lys-N, strong cation exchange
Metaproteomics of Microbial Phyllosphere Communities
Maize chloroplast membrane bundle sheath
Non-ionic Detergent Phase Extraction for the Quantitative Proteomic Analysis of Heart Membranes Prot
Q02c28
Maize chloroplast membrane mesophyll
2D reference map of *Herbaspirillum seropedicae*
Natural killer cell secretory lysosome (granule) membrane proteome
N-acetyltransferase A
Candida glabrata secreted proteins - regulation by ACE2 transcription factor
HUPO test samples
A high-density, organ-specific proteome map for Arabidopsis thaliana
Cerebrospinal fluid (CSF) proteome, Zhang et al.
Ancient fossil bone sequencing
Q02a28
Sequential Labeling for Protein Quantitation, Raijmakers et al.



Links to PRIDE BioMart Documentation

- [PRIDE BioMart Description and Help Page](#)
- [PRIDE BioMart Web Service - Accessing the Web Service](#)
- [Deploying a PRIDE BioMart Alongside Your Local PRIDE Installation](#)

EMBL-EBI  **EBI Search** All Databases [Reset ?](#) [Advanced Search](#) [Give us feedback](#)

Databases Tools EBI Groups Training Industry About Us Help [Site Index](#)  

EBI > Databases > Proteomics > PRIDE > PRIDE BioMart

Please select columns to be included in the output and hit 'Results' when ready

Select Attributes for Output

Experiment Attributes

- Project ID
- Project Name
- PRIDE Experiment Accession
- Experiment Title
- Experiment Short Label
- Contact Name
- Institution
- Contact Details (Email)
- Reference Line
- PubMed ID (CiteXplore)
- DOI (Digital Object Identifier)

Sample Attributes

- Sample Name
- Sample Description Comment
- Taxonomy Term (NEWT / NCBI Taxon)
- Taxonomy ID (NEWT / NCBI Taxon)
- Tissue Ontology Term (BRENDA)
- BRENDA ID (Tissue)
- Cell Type Term (CL)
- CL ID (Cell Type)
- Gene Ontology Term (GO)
- GO ID (Gene Ontology)
- Human Disease Term (DOID)
- DOID ID (Human Disease)

Protein Identification Attributes

- Submitted Protein Accession
- Protein Database
- Protein Database Version
- Protein Score
- Protein Threshold
- Search Engine

Dataset
PRIDE

Filters
Filter by Project : 2D reference map of *Herbaspirillum seropedicae*


Attributes
PRIDE Experiment Accession
Submitted Protein Accession



Dataset
[None Selected]

Links to PRIDE BioMart Documentation



- [PRIDE BioMart Description and Help Page](#)
- [PRIDE BioMart Web Service - Accessing the Web Service](#)
- [Deploying a PRIDE BioMart Alongside Your Local PRIDE Installation](#)

PRIDE BioMart

EMBL-EBI  **EB-eye Search** All Databases [Reset ?](#) [Advanced Search](#) [Give us feedback](#)

Databases Tools EBI Groups Training Industry About Us Help [Site Index](#)  

EBI > Databases > Proteomics > PRIDE > PRIDE BioMart

Dataset
PRIDE

Filters
Filter by Project : 2D
reference map of
Herbaspirillum seropedicae

Attributes
PRIDE Experiment Accession
Submitted Protein Accession

Dataset
[None Selected]




Export all results to Unique results only

Email notification to

View rows as Unique results only

PRIDE Experiment Accession	Submitted Protein Accession
1880	CAM32534.1
1880	CAM32674.1
1880	CAM32559.1
1880	CAM32640.1
1880	CAM32585.1
1880	CAM32629.1
1880	CAM32626.1
1880	CAM32583.1
1880	CAM32583.1
1880	CAM32547.1

biomart version 0.7

- Links to PRIDE BioMart Documentation
-  [PRIDE BioMart Description and Help Page](#)
 -  [PRIDE BioMart Web Service - Accessing the Web Service](#)
 -  [Deploying a PRIDE BioMart Alongside Your Local PRIDE Installation](#)



The spectacular bit: across-BioMart queries!

Question: “Which proteins, identified in PRIDE, in blood plasma, are transcribed from genes located in chromosome 11”

→ PRIDE
→ Ensembl

bio::mart

HOME MARTVIEW MARTSERVICE DOCS CONTACT NEWS CREDITS

New Count Results URL XML Perl Help

Dataset 1895 / 37435 Genes
Homo sapiens genes (NCBI36)
Filters
Chromosome: 11
Attributes
Ensembl Gene ID
Ensembl Transcript ID
Gene Start (bp)
Gene End (bp)
Chromosome Name
Associated Gene Name
Ensembl Protein ID

Dataset 498 / 8173 Experiments
PRIDE
Filters
Filter by Tissue : blood plasma
Attributes
PRIDE Experiment Accession
Experiment Title
Submitted Protein Accession
Uniprot Accession

Export all results to: File TSV Unique results only Go

Email notification to: [input field]

View: 10 rows as HTML Unique results only

Ensembl Gene ID	Ensembl Transcript ID	Gene Start (bp)	Gene End (bp)	Chromosome Name	Associated Gene Name	Ensembl Protein ID	PRIDE Experiment Accession	Experiment Title	Submitted Protein Accession	Uniprot Accession
ENSG00000221842	ENST00000335295	5203272	5207201	11	HBB	ENSP00000333994	13	HUPO Plasma Proteome Project, Lab # 2 Expt # 17	IPI00218816	P68871
ENSG00000118137	ENST00000236850	116211677	116213571	11	APOA1	ENSP00000236850	13	HUPO Plasma Proteome Project, Lab # 2 Expt # 17	IPI00021841	P02647
ENSG00000118137	ENST00000375320	116211677	116213571	11	APOA1	ENSP00000364469	13	HUPO Plasma Proteome Project, Lab # 2 Expt # 17	IPI00021841	P02647
ENSG00000118137	ENST00000375323	116211677	116213571	11	APOA1	ENSP00000364472	13	HUPO Plasma Proteome Project, Lab # 2 Expt # 17	IPI00021841	P02647
ENSG00000118137	ENST00000359492	116211677	116213571	11	APOA1	ENSP00000352471	13	HUPO Plasma Proteome Project, Lab # 2 Expt # 17	IPI00021841	P02647
ENSG00000180210	ENST00000311907	46697331	46717631	11	F2	ENSP00000308541	13	HUPO Plasma Proteome Project, Lab # 2 Expt # 17	IPI00019568	P00734
ENSG00000149131	ENST00000278407	57121436	57138902	11	SERPING1	ENSP00000278407	13	HUPO Plasma Proteome Project, Lab # 2 Expt # 17	IPI00291866	P05155
ENSG00000110245	ENST00000375345	116205834	116208998	11	APOC3	ENSP00000364494	13	HUPO Plasma Proteome Project, Lab # 2 Expt # 17	IPI00021857	P02656
ENSG00000110245	ENST00000227667	116205834	116208998	11	APOC3	ENSP00000227667	13	HUPO Plasma Proteome Project, Lab # 2 Expt # 17	IPI00021857	P02656
ENSG00000110169	ENST00000265983	6408858	6418830	11	HPX	ENSP00000265983	13	HUPO Plasma Proteome Project, Lab # 2 Expt # 17	IPI00022488	P02790

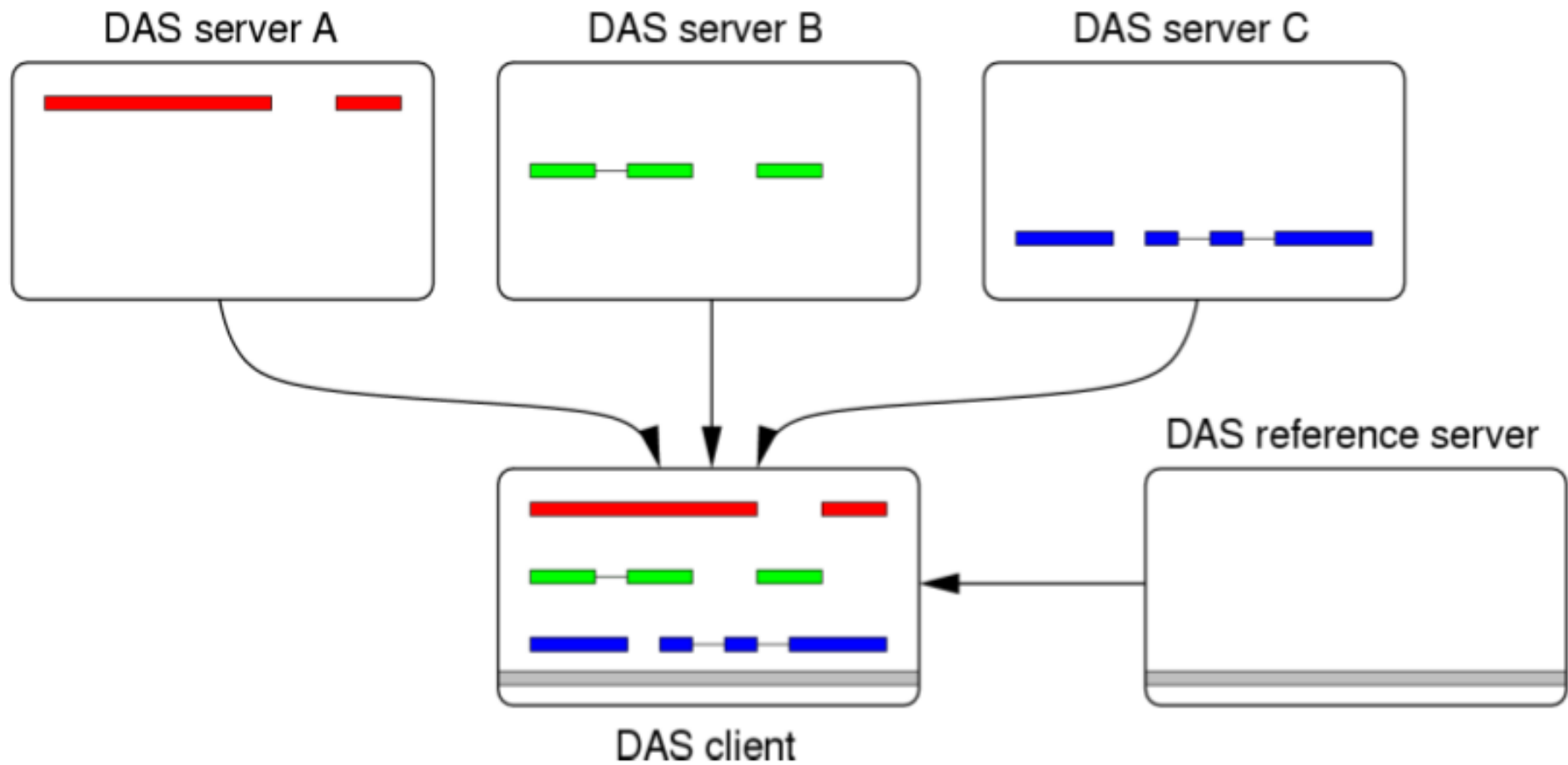
biomart version 0.7

www.biomart.org

DAS

(<http://www.ebi.ac.uk/pride-das/das/PrideDataSource>)

Distributed Annotation System



DASTY

PRIDE DAS server

```
-<DASGFF>
-<GFF href="http://www.ebi.ac.uk/pride-das/das/PrideDataSource/features?segment=P05067">
-<SEGMENT id="P05067" start="0" stop="770" version="7dd43312cd29a262acdc0517230bc5ca" label="P05067">
+<FEATURE id="P05067_coverage_1" label="protein_coverage_part_1"></FEATURE>
+<FEATURE id="P05067_coverage_2" label="protein_coverage_part_2"></FEATURE>
+<FEATURE id="P05067_coverage_3" label="protein_coverage_part_3"></FEATURE>
+<FEATURE id="P05067_coverage_4" label="protein_coverage_part_4"></FEATURE>
+<FEATURE id="P05067_coverage_5" label="protein_coverage_part_5"></FEATURE>
+<FEATURE id="P05067_coverage_6" label="protein_coverage_part_6"></FEATURE>
+<FEATURE id="P05067_coverage_7" label="protein_coverage_part_7"></FEATURE>
+<FEATURE id="P05067_coverage_8" label="protein_coverage_part_8"></FEATURE>
+<FEATURE id="P05067_coverage_9" label="protein_coverage_part_9"></FEATURE>
+<FEATURE id="P05067_coverage_10" label="protein_coverage_part_10"></FEATURE>
+<FEATURE id="P05067_coverage" label="protein_coverage"></FEATURE>
-<FEATURE id="LEVPTDGNAGLLAEPQIAMFCGR_18_40" label="LEVPTDGNAGLLAEPQIAMFCGR">
  <TYPE id="Polypeptide" cvId="SO:0000358" category="3">Polypeptide</TYPE>
  -<METHOD id="PRIDE" cvId="ECO:0000160">
    Inferred from protein separation followed by fragment identification
  </METHOD>
  <START>18</START>
  <END>40</END>
  <SCORE>3.0</SCORE>
  <LINK href="http://www.ebi.ac.uk/pride/biomart/martview?VIRTUALSCHEMANAME=default&ATTRIBUTES=pride.default.pride_attributes.peptide_spectrum_ref&
  FILTERS=pride.default.pride_filters.peptide_sequence."LEVPTDGNAGLLAEPQIAMFCGR"|pride.default.pride_filters.uniprot_ac."P05067"&VISIBLEPANEL=resultspanel">View mass
  spectra</LINK>
  <LINK href="http://www.ebi.ac.uk/pride/biomart/martview?VIRTUALSCHEMANAME=default&
  ATTRIBUTES=pride.default.pride_attributes.reference_line|pride.default.pride_attributes.pubmed_id|pride.default.pride_attributes.doi&
  FILTERS=pride.default.pride_filters.peptide_sequence."LEVPTDGNAGLLAEPQIAMFCGR"|pride.default.pride_filters.uniprot_ac."P05067"&VISIBLEPANEL=resultspanel">View
  literature</LINK>
  <LINK href="http://www.ebi.ac.uk/pride/biomart/martview?VIRTUALSCHEMANAME=default&
  ATTRIBUTES=pride.default.pride_attributes.experiment_ac|pride.default.pride_attributes.experiment_title|pride.default.pride_attributes.experiment_short_title|pride.default.pride_attributes.proj
  FILTERS=pride.default.pride_filters.peptide_sequence."LEVPTDGNAGLLAEPQIAMFCGR"|pride.default.pride_filters.uniprot_ac."P05067"&VISIBLEPANEL=resultspanel">View list of
  experiments</LINK>
  <LINK href="http://www.ebi.ac.uk/pride/biomart/martview?VIRTUALSCHEMANAME=default&
  ATTRIBUTES=pride.default.pride_attributes.sample_name|pride.default.pride_attributes.sample_description_comment|pride.default.pride_attributes.newt_name|pride.default.pride_attributes.new
  FILTERS=pride.default.pride_filters.peptide_sequence."LEVPTDGNAGLLAEPQIAMFCGR"|pride.default.pride_filters.uniprot_ac."P05067"&VISIBLEPANEL=resultspanel">View sample
  details</LINK>
</FEATURE>
-<FEATURE id="LNMHMNVQNGK_41_51" label="LNMHMNVQNGK">
```


DASTY3 client

(<http://wwwdev.ebi.ac.uk/dasty/pride/demo/dasty.html>)

DASTY client

- Dasty, a web client which retrieves, integrates and visualizes protein annotations.
- It collects data from different DAS sources and then merge it to provide the user with a unified view of the sequence-annotated features.

PRIDE DAS server: Dasty example (1)

(<http://wwwdev.ebi.ac.uk/dasty/pride/demo/dasty.html>)

EBI - DASTY

Search

Protein ID: Registry label:

Examples: [P03067](#), [P03973](#), [P13569](#), [MDM2](#), [MOUSE](#), [BRCA1](#), [HUMAN](#)...

58 annotations retrieved in 01 seconds, 262 milliseconds 116 annotations retrieved in 01 seconds, 000 milliseconds

System Information

100% - Dasty loaded all features


Done Searching [Annotation Servers Log](#)

Sequence

Sequence ID: [P03973](#)
Sequence length: 132
Format:

```
>P03973 132 bp
RSISLRFVLLVLAIGTLPFNAVGGSGKFEADYCFKKSQAQCLYKXKFCQEDWQCPK
KRCCFDTGCIKCLDPVDTFPIFRKFKGKCVTYGQCLMLNFFNFCMDGQCKRLKRCMG
MGKSCVSPVKA
```

3D Structure



Positional Features

Manipulation Options:

FEATURE TYPE	METHOD	LABELS	FEATURE ANNOTATIONS	EXPERIMENT
signal_peptide	UniProt	LNIPROTKB_P03973		Inferred by curator (ECO:000000)
maturation_peptide	UniProt	LNIPROTKB_P03973		Inferred by curator (ECO:000000)
signal_peptide	UniProt	WAP_1 WAP_2 		Inferred by curator (ECO:000000)
signal_peptide_domain	UniProt	WAP_1 WAP_2 		Inferred by curator (ECO:000000)
signal_peptide_region	UniProt	Trypsin inhibitory dom		Inferred by curator (ECO:000000)
signal_peptide_region	UniProt	Reactive bond for try		Inferred by curator (ECO:000000)
disulfide_crosslinked	UniProt	LNIPROTKB_P03973		Inferred by curator (ECO:000000)
disulfide_crosslinked	UniProt	LNIPROTKB_P03973		Inferred by curator (ECO:000000)
disulfide_crosslinked	UniProt	LNIPROTKB_P03973		Inferred by curator (ECO:000000)
disulfide_crosslinked	UniProt	LNIPROTKB_P03973		Inferred by curator (ECO:000000)
beta_strand	UniProt	LNIPROTKB_P03973		Inferred by curator (ECO:000000)
consensus_region	Inferred from electron/protein_coverage			coverage
polyptide	Inferred from protein sMKSSGLPFLVLLAL			9300
polyptide	Inferred from protein sYKPFEGQSDWQCP			405
polyptide	Inferred from protein sCLDPVDTFPIFR			9187
polyptide	Inferred from protein sCLDPVDTFPIFR			9332
polyptide	Inferred from protein sDLKDDMGK			0667
polyptide	Inferred from protein sCVSPVKA			452

The annotation is in accordance with the version of the protein sequence.
 Caution! The annotation may refer to an old version of the protein sequence, so the position of features may be incorrect.
 Group of features classified by the annotation server.
 Features grouped in the same line by Dasty2.

Non Positional Features

PRIDE DAS server: Dasty example (2)

(<http://wwwdev.ebi.ac.uk/dasty/pride/demo/dasty.html>)

EBI > DASTY

Search

Protein ID: P03973 [Registry label: BioSapiens](#)

Examples: P05067, P03973, P13569, MDM2 MOUSE, BRCA1 HUMAN...

58 annotations retrieved in 01 seconds, 262 milliseconds

System Information

Sequence

Sequence ID: P03973
Sequence length: 132
Format: FASTA


>P03973 132 bp
KSSGLFPFLVLLALGTLAPNAVEGSGKSFKAGVCPFKKSAQCLRYKKPEQSDWQCPGK
KRCCPDTGIGKCLDPVDTNPTRRPGKCPVTYGQCIMLNPNFCMDQCKRDLKCCMG
MGKSCVSPVKA

CLDPVDTNPTRR_72_84

Feature ID : CLDPVDTNPTRR_72_84
Feature Label : CLDPVDTNPTRR
Type : Polypeptide
Type ID : Polypeptide
Category : 9332
Method : Inferred from protein separation followed by fragment identification
Start : 72
End : 84
Score : 1.0
Orientation :
Phase :
Links :
[View literature](#)
[View mass spectra](#)
[View list of experiments](#)
[View sample details](#)

3D Structure

2Z7F.1



Positional Features

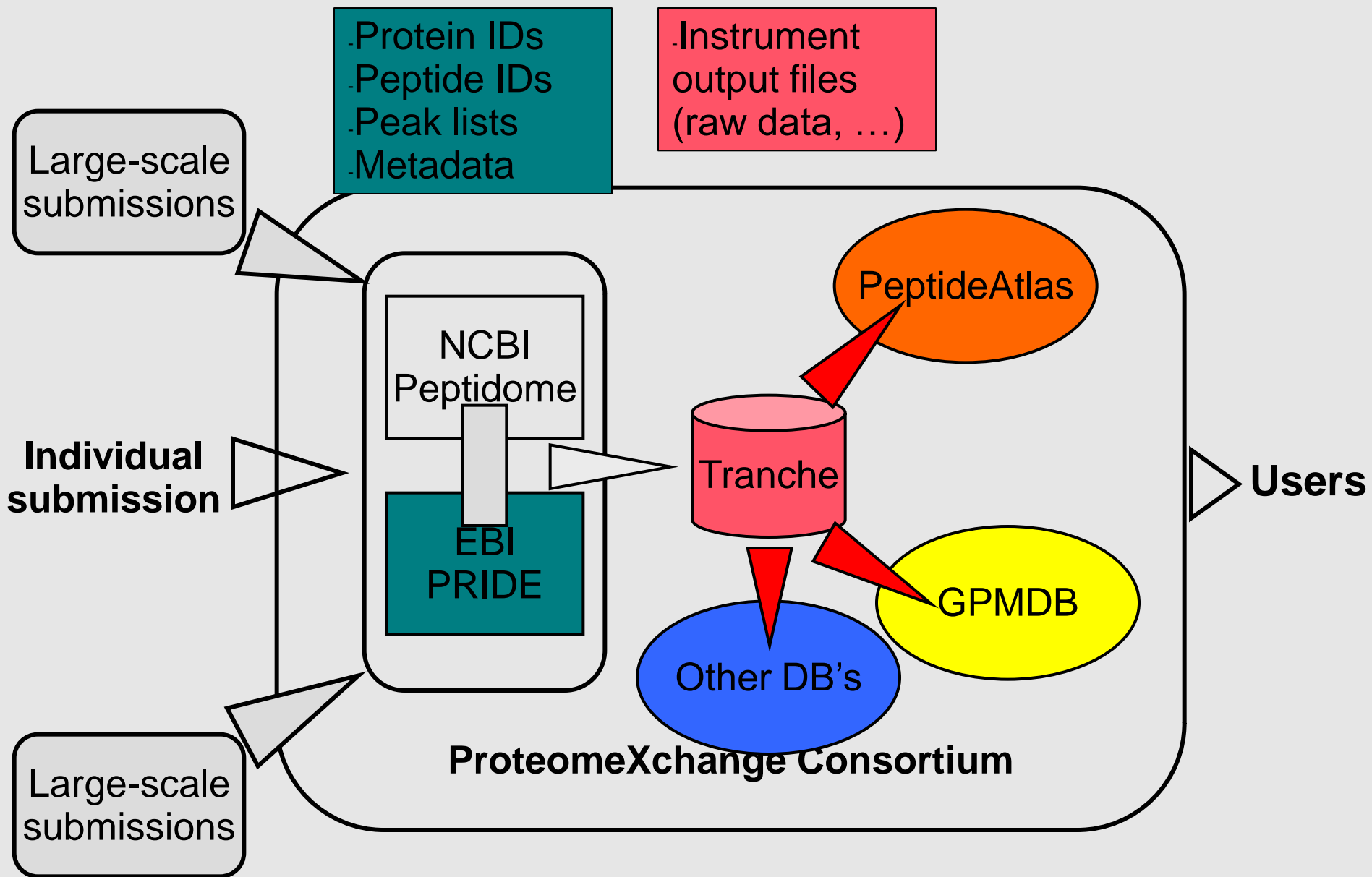
Manipulation Options:

FEATURE TYPE	FEATURE ANNOTATIONS	SERVER NAME	EXPERIMENT
signal peptide		uniprot	inferred by curator (ECO:000000)
mature protein region		uniprot	inferred by curator (ECO:000000)
polypeptide domain		uniprot	inferred by curator (ECO:000000)
polypeptide region		uniprot	inferred by curator (ECO:000000)
polypeptide region		uniprot	inferred by curator (ECO:000000)
disulfide crosslinked n		uniprot	inferred by curator (ECO:000000)
disulfide crosslinked n		uniprot	inferred by curator (ECO:000000)
disulfide crosslinked n		uniprot	inferred by curator (ECO:000000)
disulfide crosslinked n		uniprot	inferred by curator (ECO:000000)
beta strand		uniprot	inferred by curator (ECO:000000)
consensus region		pride.das 1.6	coverage
Polypeptide		pride.das 1.6	9300
Polypeptide		pride.das 1.6	485
Polypeptide		pride.das 1.6	8187
Polypeptide		pride.das 1.6	9332
Polypeptide		pride.das 1.6	8667
Polypeptide		pride.das 1.6	452

Jmol

PRIDE AND OTHER REPOSITORIES: ProteomeXchange





The PRIDE Team

Attila Csordas



Juan Vizcaino



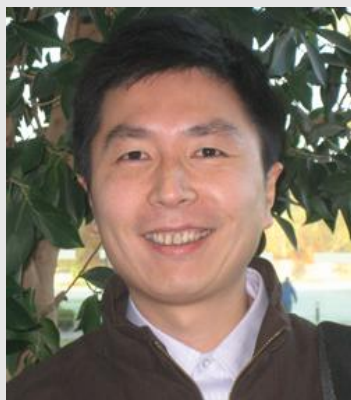
Richard Côté



Florian Reisinger



Henning Hermjakob



Rui Wang



Joe Foster
(Ph.D. student)



Andreas Schonegger
(Trainee)

Special thanks, collaborations and funding

Rafael Jimenez



Jose Villaveces



dani@ebi.ac.uk
pride-support@ebi.ac.uk

PRIDE collaborators



Funding



Thank you!

Questions?