

Pedant-Pro™ Sequence Analysis Suite Release Notes

Pedant-Pro version 3.3

Biomax Informatics AG



Pedant-Pro™ Sequence Analysis Suite Release Notes: Pedant-Pro version 3.3

Biomax Informatics AG



PEDANTRLN0912

Publication date December 2009

Copyright © 2006-2009 Biomax Informatics AG

Biomax Informatics AG; Sitz: Planegg bei München, Amtsgericht München, HRB 134442; Ust-IdNr. DE 191 603 142; Steuer-Nr.: 143 / 100 / 10945; Vorstand: Dr. Klaus Heumann; Vorsitzender des Aufsichtsrates: Prof. Dr. Hans-Werner Mewes

Table of Contents

1. Release Notes Pedant-Pro version 3.3	1
Upgrade information	1
Version information	1
Upgrade and patch information	1
Supported client platforms and operating systems	1
Supported server platforms and operating systems	1
Supported relational database management systems (RDBMS)	2
Available installation modes	2
Software modules	2
Database schema	2
Software description	3
New in Pedant-Pro version 3.3	3
Overview	3
New methods	4
Modified methods	4
New GUI features	4
New main service features	13
2. Contact Biomax	14

List of Figures

1.1. Pedant-Pro GUI new look and feel	5
1.2. Gene structure list	6
1.3. Protein length distribution list	6
1.4. Protein molecular weight distribution list	6
1.5. Protein isoelectric point distribution list	7
1.6. BLAST overview list	7
1.7. Pathway coverage - result list	8
1.8. Pathway coverage - individual pathway overview page	8
1.9. Access to genome comparison results from a gene list	9
1.10. DNA Viewer - Overview	9
1.11. DNA Viewer — "Navigation and settings" panel	10
1.12. DNA Viewer -"Sequence" panel	10
1.13. Protein Viewer - Overview	11
1.14. Protein Viewer — "Navigation and settings" panel	11
1.15. Export from a report page	12
1.16. Export from a gene list	12
1.17. Export from a sequence page	12
1.18. Export from a method results page	12

List of Tables

1.1. Naming of exported files 12

Chapter 1. Release Notes Pedant-Pro version 3.3

Upgrade information

Version information

These release notes are current as of December 2009.

This document covers all new features of the release version 3.3 of the Pedant-Pro Sequence Analysis Suite. For all other release information pertaining to this version, see the release notes for the previous releases.

Upgrade and patch information

Version 3.3 of the Pedant-Pro Sequence Analysis Suite is a full public release.

Supported client platforms and operating systems

The Pedant-Pro Sequence Analysis Suite version 3.3 client supports the following operating systems:

- Microsoft® Windows® 2000
- Microsoft® Windows® XP
- Microsoft® Windows® Vista®

The client system supports the following browsers

- Microsoft Internet Explorer® 6.0, 7
- Mozilla Firefox® 2.x, 3.x

Note

The client software might perform adequately using other browsers on other platforms (e.g., Linux® systems) although this performance has not been tested and is not supported by Biomax.

The client system supports the following Java™ Virtual Machine (JVM):

- JVM version 5.0
- To use the Biomax Pathway Viewer, a Java Web Start installation, which is included in Java version Java 5, is required on the client system.

Supported server platforms and operating systems

The following platforms and operating systems are supported by Pedant-Pro version 3.3:

- SuSE® Linux Enterprise Server 9, 10 (x86, x86_64)

Note

The server software may perform adequately on other operating systems (e.g., Red Hat® Linux or Ubuntu Linux), although this performance has not been tested and is not supported by Biomax.

Supported relational database management systems (RDBMS)

The Pedant-Pro Sequence Analysis Suite version 3.3 supports the following RDBMS:

- Sun MySQL™ 5.0 and MySQL 5.1
- Oracle Database 10g

Available installation modes

The following installation modes are available:

- In-house installation — the Pedant-Pro server is installed at the customer's site. Maintenance and administration can be provided as a service by Biomax, but can also be performed by the customer. Support is provided by Biomax according to a contractual agreement.
- Hosted service — the Pedant-Pro server is located at the Biomax facility and the customer accesses the Web server via the Internet. Maintenance and administration are provided by Biomax. Support is provided by Biomax according to a contractual agreement.

Note

The scripting interfaces (PCLI and PSSI) require direct server access and are not available for hosted services installations.

Software modules

The following modules are available in Pedant-Pro version 3.3:

- Pedant-Pro service: version R-3_72
- Pedant-Pro graphical user interface (GUI): version R-3.2.29
- Server-side viewers (SSV): version R-4_36
- Cluster Viewer: version R-2_0_13
- PCLI: version R-3_41
- PSSI: version R-3_72
- Annotation module: version R-3_35
- Retrieval Manager version R-1_12

Database schema

Pedant-Pro version 3.3 uses the database schema 3.6.1

Software description

The Pedant-Pro Sequence Analysis Suite provides fast, comprehensive, enterprise-scale annotation of individual sequences and complete genomes. The initial automatic data analysis can be supplemented by manual annotation conveniently. The Pedant-Pro database represents a company's sequence, gene and protein repository with extensive information from public and proprietary data sources for each coding entity. The workflow architecture of the Pedant-Pro suite provides maximum flexibility for current and future requirements in research and development. Comfortable user interfaces guarantee optimal usability and various interfaces permit seamless integration into large automated computing environments. The Pedant-Pro system includes the following features:

- Predicts genes and open reading frames (ORFs) for prokaryotic and eukaryotic organisms using the most reliable computational methods.
- Detects numerous types of non-coding genes (e.g., rRNAs, tRNAs) using a proprietary Biomax database that contains more than 1.2 million entries
- Performs a variety of similarity searches to predict potential protein functions and protein structure: protein domains (e.g., Pfam domains) are listed, assignment to several functional ontologies (e.g., Enzyme Commission (EC) numbers and Gene Ontology (GO)) is performed, and two-dimensional and three-dimensional characteristics are displayed
- Provides comfortable user interfaces for exploring the results: users can explore potential protein functionality, use sophisticated viewers to display DNA and protein structure, and easily display gene models of the genome; pathways are displayed for a single species or in comparison with other taxa; families of paralog proteins are displayed with domains and potential functions
- Provides manual text annotation capability (even in distributed systems)
- Provides a comprehensive report page for each gene and protein with references to public and proprietary databases
- Provides a highly dynamic system based on workflow architecture: the Pedant-Pro system comes with pre-configured workflows defined by Biomax scientists; the flexible architecture allows definition of company-specific workflows, which cover all required use cases
- Provides powerful scripting interfaces, the Pedant Service Scripting Interface (PSSI) and the Pedant-Pro command-line interface (PCLI), for integration of the Pedant-Pro suite into large computing environments
- Supports high-throughput analysis with industrial-scale computing grids
- Provides a powerful search interface for searching sequence, pattern, identifiers and text; the text search offers the full functionality of the BioRS Integration and Retrieval System text-search capabilities

New in Pedant-Pro version 3.3

Overview

Version 3.3 of the Pedant-Pro Sequence Analysis Suite adds a number of features improving functionality and ease of use for both users and administrators:

- **Additional and improved methods for gene prediction and automatic annotation**

- **New overview tables for contigs, genes, proteins, pathways and BLAST results**
- **Improved DNA and Protein Viewers**
- **Improved sorting of contig tables**
- **Improved search functionality**
- **Improved GUI and Command Line Interface data export functionality**
- **Additional access points for genome comparison**
- **New Pedant-Pro look and feel**

This document describes these features in detail as well as numerous smaller additions and extensions introduced with Pedant-Pro version 3.3.

New methods

Pedant-Pro 3.3 supports now the following methods:

- **Glimmer3**: gene finding in microbial DNA, especially the genomes of bacteria, archaea, and viruses.
- **GeneMark**: gene prediction in bacteria (requires a commercial license).
- **Biomax combined bacterial gene prediction**: the method "combprokbp" performs a combined gene prediction for bacterial genomes combining ab initio gene prediction (e.g. with Glimmer) with additional extrinsic evidences from reliable proteins of a related organism in the public domain.
- **TransTermHP**: finds Rho-independent transcription terminators in bacterial genomes.
- **Blastgeclustalw**: supports the generation of multiple alignments from blastp results.
- **Blasthitgetitle**: configurable method that allows the automatic assignment of meaningful titles/descriptions for genes based on blastp analysis against the UniProt database.
- **Pathwaycoverage**: allows to determine the coverage of pathways (e.g MetaCat pathways) in an analysis by examining the identified EC numbers.

Modified methods

- **Pfam**: the pfam method now supports HMMER3. The Pfam database provides from release 24.0 only HMMER3 HMMs which are not backward compatible to HMMER2.
- **Improved prediction of EC numbers**: All "blastp" methods now calculate a weightedAverageScore and a weightedIdentity value for the BLAST results. The first is a DB size independent and comparable quality measure for a BLAST result taking the query and subject coverage into account. The second is a heuristic approach to overcome the BLAST bias towards long hits, which is especially important for enzyme prediction (meanwhile the BLAST bias has been fixed in the NCBI tool kit).

New GUI features

All pages of the Pedant-Pro version 3.3 GUI have been updated to a new look and feel.

Figure 1.1. Pedant-Pro GUI new look and feel

The screenshot shows the Pedant-Pro GUI interface. On the left is a navigation tree with categories like 'Contigs', 'Genes and Genetic Elements', 'Protein Encoding Genes', 'Genetic Elements', 'Group', 'Protein Function', 'BLAST', 'Protein Structure', and 'Taxonomy'. The main window displays a 'List of Genes' table with 1862 entries. The table has columns: Sel, Gene, Gene name, Contig, Start, Stop, Description, Best BLAST Hit, and Hit ID. The first few rows are as follows:

Sel	Gene	Gene name	Contig	Start	Stop	Description	Best BLAST Hit	Hit ID
<input type="checkbox"/>	LB40001	NC_008614	31	1395	chromosomal replication initiation protein	P. DNA-directed DNA replication initiation protein - Lactobacillus acidophilus ATCC 4796	UNIPROT_C2HKC4	
<input type="checkbox"/>	LB40002	NC_008614	1576	2709	DNA polymerase III, beta chain	P. DNA polymerase III, beta chain EC=2.7.7.7 - Lactobacillus acidophilus ATCC 4796	UNIPROT_C2HKC9	
<input type="checkbox"/>	LB40003	NC_008614	2925	3152	hypothetical protein	P. Putative uncharacterized protein - Lactobacillus acidophilus	UNIPROT_O5FN13	
<input type="checkbox"/>	LB40004	NC_008614	3155	4279	recombination protein F	P. Recombination protein F - Lactobacillus acidophilus ATCC 4796	UNIPROT_C2HKX1	
<input type="checkbox"/>	LB40005	NC_008614	4283	6244	DNA gyrase subunit B	P. DNA gyrase subunit B EC=5.99.1.3 - Lactobacillus acidophilus ATCC 4796	UNIPROT_C2HKX9	
<input type="checkbox"/>	LB40006	NC_008614	6258	8729	DNA gyrase subunit A	P. DNA gyrase subunit A EC=5.99.1.3 - Lactobacillus acidophilus ATCC 4796	UNIPROT_C2HKJ9	
<input type="checkbox"/>	LB40007	NC_008614	8933	9226	30S ribosomal protein S6	P. 30S ribosomal protein S6 - Lactobacillus acidophilus ATCC 4796	UNIPROT_C2HKJR	
<input type="checkbox"/>	LB40008	NC_008614	9275	9790	single-stranded DNA-binding protein	P. Single-stranded DNA-binding protein - Lactobacillus acidophilus ATCC 4796	UNIPROT_C2HKJ7	
<input type="checkbox"/>	LB40009	NC_008614	9822	10055	30S ribosomal protein S18	P. 30S ribosomal protein S18 - Lactobacillus ulhuenensis DSM 16047	UNIPROT_C2HKV2	
<input type="checkbox"/>	LB40011	NC_008614	10237	12255	putative signaling protein	P. Orib family phosphatase - Lactobacillus acidophilus ATCC 4796	UNIPROT_C2HKJ5	
<input type="checkbox"/>	LB40012	NC_008614	12271	12723	50S ribosomal protein L9	P. 50S ribosomal protein L9 - Lactobacillus acidophilus ATCC 4796	UNIPROT_C2HKJ4	
<input type="checkbox"/>	LB40013	NC_008614	12752	14143	replicative DNA helicase	P. Replicative DNA helicase - Lactobacillus acidophilus ATCC 4796	UNIPROT_C2HKJ3	
<input type="checkbox"/>	LB40014	NC_008614	14388	15326	putative allylphosphonate ABC transporter	P. Putative allylphosphonate ABC transporter - Lactobacillus acidophilus	UNIPROT_O5FN09	
<input type="checkbox"/>	LB40015	NC_008614	15433	16359	hypothetical protein	P. Orib family protein - Lactobacillus acidophilus ATCC 4796	UNIPROT_C2HKJ1	
<input type="checkbox"/>	LB40016	NC_008614	16483	17358	fructokinase	P. Fructokinase EC=2.7.1.4 - Lactobacillus acidophilus ATCC 4796	UNIPROT_C2HKJ0	
<input type="checkbox"/>	LB40017	NC_008614	17507	17692	putative general stress response	P. Putative uncharacterized protein - Lactobacillus acidophilus	UNIPROT_O5EM29	
<input type="checkbox"/>	LB40018	NC_008614	18036	17795	hypothetical protein	P. Putative uncharacterized protein - Lactobacillus acidophilus	UNIPROT_O5EM29	

Contig overview tables

For contigs additional overview tables are available:

- **Contig Information:** Additional information about contigs in the analysis, e.g. base composition, length.
- **List of contigs with genes:** List of all contigs in the analysis that contain protein encoding genes

Improved sorting of contig tables

Contig lists can now be sorted by additional contig features:

- contig length
- number of encoded genes

Gene and protein overview tables

New overview lists have been added to the "Genes and genetic elements" item in the navigation tree of the analysis window. Dependent on the analysis not all lists might be available.

Genes by gene structure

The genes by gene structure list allows access to groups of genes based on their gene structure. The gene structure list is only available for analyses of eukaryotic organisms where genes are composed of exons and introns.

Figure 1.2. Gene structure list

Gene Structure		
No. of Proteins	% Proteins	No. of Exons
4571	36.9%	Genes with 1 Exon
3284	26.5%	Genes with 2 Exons
2198	17.8%	Genes with 3 Exons
1172	9.5%	Genes with 4 Exons
601	4.9%	Genes with 5 Exons
312	2.5%	Genes with 6 Exons
165	1.3%	Genes with 7 Exons
61	0.5%	Genes with 8 Exons
46	0.4%	Genes with 9 Exons
21	0.2%	Genes with 10 Exons
17	0.1%	Genes with 11 Exons
5	0.0%	Genes with 12 Exons
2	0.0%	Genes with 13 Exons
5	0.0%	Genes with 14 Exons
5	0.0%	Genes with ≥15 exons

Genes by protein length distribution

The protein length distribution list allows access to groups of genes based on the length of their corresponding protein.

Figure 1.3. Protein length distribution list

Protein Length Distribution		
No. of Proteins	% Proteins	Length
51	2.7%	Proteins < 50 aa
348	18.7%	Proteins 50-149 aa
682	36.6%	Proteins 150-299 aa
755	40.5%	Proteins 300-999 aa
25	1.3%	Proteins 1000-3000 aa
1	0.1%	Proteins > 3000 aa

Genes by protein molecular weight distribution

The protein molecular weight distribution list allows access to groups of genes based on the molecular weight of their corresponding protein.

Figure 1.4. Protein molecular weight distribution list

Protein Molweight Distribution		
No. of Proteins	% Proteins	Molweight
34	1.8%	Proteins ≤ 5.0E3 Da
172	9.2%	Proteins ≤ 1.0E4 Da
496	26.6%	Proteins ≤ 2.0E4 Da
1509	81.0%	Proteins ≤ 5.0E4 Da
1823	97.9%	Proteins ≤ 1.0E5 Da
39	2.1%	Proteins > 1.0E5 Da
5	0.3%	Proteins > 2.0E5 Da
0	0%	Proteins > 5.0E5 Da

Genes by protein isoelectric point distribution

The protein isoelectric point distribution list allows access to groups of genes based on the isoelectric point of their corresponding protein.

Figure 1.5. Protein isoelectric point distribution list

Protein Isoelectric Point Distribution					
No. of Proteins	% Proteins	Isoelectric Point			
2	0.1%	Proteins $pl < 3.5$	very strong acidic		
72	3.9%	Proteins $3.5 \leq pl < 4.5$	strong acidic		
417	22.4%	Proteins $4.5 \leq pl < 5.5$	acidic		
238	12.8%	Proteins $5.5 \leq pl < 6.5$	weak acidic		
165	8.9%	Proteins $6.5 \leq pl < 7.5$	neutral		
98	5.3%	Proteins $7.5 \leq pl < 8.5$	weak basic		
128	6.9%	Proteins $8.5 \leq pl < 9.5$	basic		
460	24.7%	Proteins $9.5 \leq pl < 10.5$	strong basic		
282	15.1%	Proteins $pl \geq 10.5$	very strong basic		

BLAST overview lists

BLAST overview lists allow access to list of genes based on *blast* percentage of identity ranges against entries of the respective BLAST target database. Protein are grouped by the *blast* identity values of the respective type of *blast* analysis.

Figure 1.6. BLAST overview list

Best protein match (blastp UniProt)			
Proteins grouped by blast identity values			
No. of Proteins	% Proteins	% Identity	
1747	93.8%	100%	
1809	97.2%	> 90%	
1816	97.5%	≥ 50%	
1816	97.5%	> 30%	
1816	97.5%	≥ 0%	
46	2.5%	0%	

Pathway overview tables

The Pathway coverage (MetaCat) result list summarizes the analysis results for metabolic pathways based on the Biomax MetaCat ontology. For all pathways the number of identified enzymes in relation to existing number of enzymes per pathway is returned. The completeness of pathways is indicated by colors.

Figure 1.7. Pathway coverage - result list

Pathways			
Pathway coverage (MetaCat)		Automatic Pathways (UniProt)	
464 entries found			
SORT BY	▲ MetaCat Numbers	Enzyme Coverage	MetaCat Descriptions
Sorting	Metabolism	Energy	
01.01.01	2/2		assimilation of ammonia by ATP-dependent amidation
01.01.01.05	1/1		ammonia assimilation to glutamine
01.01.01.10	1/1		ammonia assimilation to NAD
01.01.02	6/7		assimilation of ammonia by reductive amination
01.01.02.01	1/2		ammonia assimilation to glutamate (reductive amination of oxoglutarate, NAD)
01.01.02.02	2/2		ammonia assimilation to glutamate (reductive amination of oxoglutarate, NADP)
01.01.02.05	1/1		ammonia assimilation to alanine (reductive amination of pyruvate)
01.01.02.20	3/3		ammonia assimilation to glycine (glycine synthase complex)
01.01.05	2/4		biosynthesis of proline and hydroxyproline

For each individual pathway an overview page is presented summarizing the enzymes identified for the selected pathway and indicating which enzymes have not been identified in the dataset. For all identified enzymes links to the respective genes encoding these enzymes are available.

Figure 1.8. Pathway coverage - individual pathway overview page

Pathway coverage (MetaCat)			
01.01 amino-acid metabolism			
01.01.20 biosynthesis of aspartate, asparagine and beta-alanine			
01.01.20.01 aspartate cycle (cytosol)			
6.3.4.5	-	argininosuccinate synthase	
4.3.2.1	-	argininosuccinate lyase	
4.2.1.2	+	fumarate hydratase	LBA0907
1.1.1.37	+	malate dehydrogenase	LBA0910 LBA0199 LBA0271
2.6.1.1	+	aspartate transaminase	LBA1516 LBA0937 LBA1695 LBA0856

Additional access points to genome comparison results

"Genes in genome comparison" gene list

A new gene list "Genes in genome comparison" has been added under the "Genes and genetic elements" item in the navigation tree of the analysis window. The list is only available for analyses where genome comparison analysis has been performed.

Access to genome comparison results from gene and Blast lists

The presence of the "🔍" icon in the GeCo column in gene and blast lists indicates that results based on comparison to another genome(s) are available. Clicking the icon opens the genome comparison report for the selected gene.

Figure 1.9. Access to genome comparison results from a gene list

List of Genes								
Order by Position		Manual Order						
Code	GeCo	Contig	Start	Stop	Description	Best BLAST Hit	Hit ID	E-value
LBA0001	NC_006814	31	1395	chromosomal replication initiation protein	P. DNA-directed DNA replication initiator protein - Lactobacillus acidophilus ATCC 4796	UNIPROT:G2HKK4_LACAC	0.0	
LBA0002	NC_006814	1576	2703	DNA polymerase III, beta chain	P. DNA polymerase III, beta chain EC=2.7.7.7 - Lactobacillus acidophilus ATCC 4796	UNIPROT:G2HKK3_LACAC	0.0	
LBA0003	NC_006814	2925	3152	hypothetical protein	P. Putative uncharacterized protein - Lactobacillus acidophilus	UNIPROT:G5FN13_LACAC	1.9E-32	
LBA0004	NC_006814	3155	4279	recombination protein F	P. Recombination protein F - Lactobacillus acidophilus ATCC 4796	UNIPROT:G2HKK1_LACAC	0.0	
LBA0005	NC_006814	4283	6244	DNA gyrase subunit B	P. DNA gyrase subunit B EC=5.99.1.3 - Lactobacillus acidophilus ATCC 4796	UNIPROT:G2HKK0_LACAC	0.0	
LBA0006	NC_006814	6258	8729	DNA gyrase subunit A	P. DNA gyrase subunit A EC=5.99.1.3 - Lactobacillus acidophilus ATCC 4796	UNIPROT:G2HKJ9_LACAC	0.0	
LBA0007	NC_006814	8933	9226	30S ribosomal protein S6	P. 30S ribosomal protein S6 - Lactobacillus acidophilus ATCC 4796	UNIPROT:G2HKJ8_LACAC	1.9E-40	

Improved DNA Viewer

The DNA Viewer of Pedant-Pro 3.3 has a new look and feel and a number of new features and changed default configurations improving functionality and ease of use.

Figure 1.10. DNA Viewer - Overview

The screenshot displays the DNA Viewer interface. At the top, there are controls for the displayed sequence range (From: 30501 bp, To: 80500 bp) and zoom settings (50% zoom in/out). Below these are scroll and options panels. The main area shows a genomic map with gene annotations (red arrows) and a detailed view of the DNA sequence (ATGTAACCTT...GAGATATCGA) with a genetic code legend for Bacterial and Plant Plastid. The interface includes a 'Show sequence from' field and buttons for 'Show', 'Translate', and 'Mark selected'.

Improved "Navigation and settings" panel





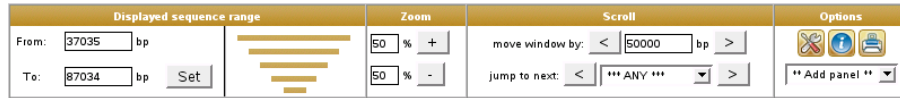
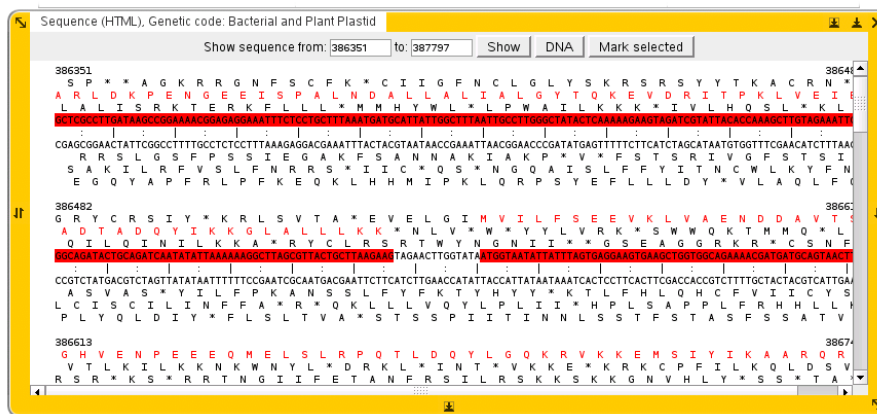
- The "Navigation and settings" panel has been restructured.
- Sequence range bars , allow to set the displayed sequence range to defined values of 150,000, 50,000, 10,000, 5,000, or 1,000 bases.
- Preferences, Help, and Printing options are available via the , , and  icons, respectively.



Figure 1.11. DNA Viewer — "Navigation and settings" panel

Improved "Sequence" panel

The "Sequence" panel is now independent from other panels and can be freely moved in horizontal direction over the other panels of the DNA Viewer. This supports a more flexible working with the different DNA Viewer panels.

Figure 1.12. DNA Viewer - "Sequence" panel

"Center selected element" button

The "Center selected element" button  in the "Contig structure" panel is available when a gene or genetic element has been selected. On click of the  button the selected element is centered on the screen

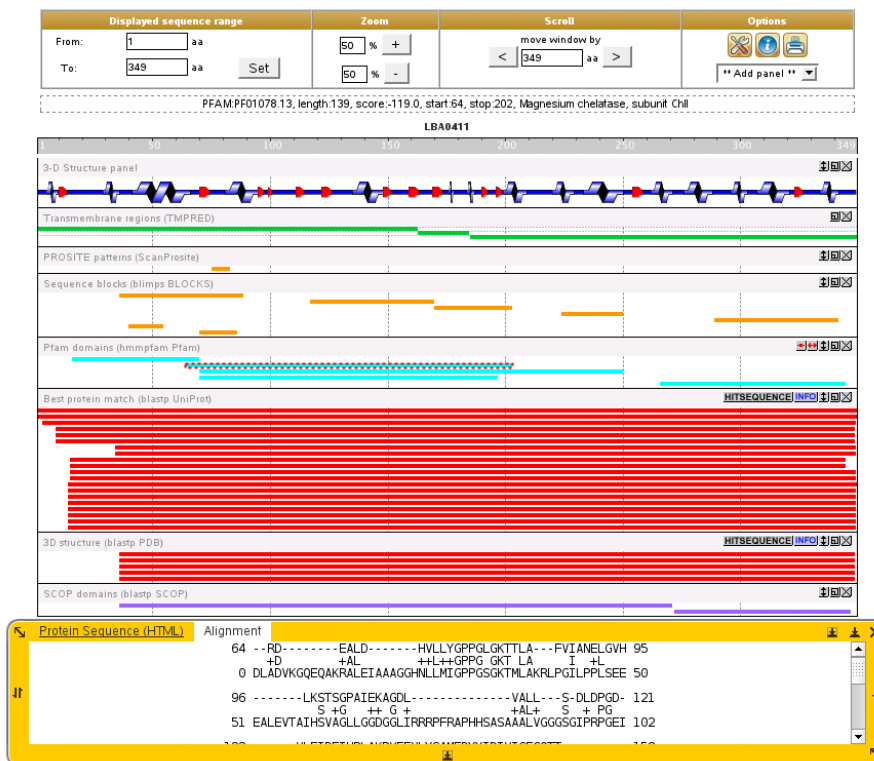
Improved orientation

By opening the DNA Viewer from a gene or genetic element report page the respective element is automatically selected and centered on the screen. When the "jump to next element" functionality of the Scroll section in the "Navigation and settings" panel is used, the next element is automatically selected and centered to the screen.

Improved Protein Viewer

The Protein Viewer of Pedant-Pro 3.3 has a new look and feel and a number of new features and changed default configurations improving functionality and ease of use.

Figure 1.13. Protein Viewer - Overview



Improved "Navigation and settings" panel




- The "Navigation and settings" panel has been restructured.
- Preferences, Help, and Printing options are available via the , , and  icons, respectively.

Figure 1.14. Protein Viewer — "Navigation and settings" panel



Improved "Sequence" panel

The sequence panel is now independent from other panels and can be freely moved in horizontal direction over the other panels of the Protein Viewer. This supports a more flexible working with the different Protein Viewer panels..

Improved search functionality

The "Gene and genetic element search" option supports now gene and genetic element description search.

Improved GUI data export

A number of improvements in respect to ease of use and defined export formats have been implemented to support export of data via the Pedant-Pro GUI.

Context dependent export

The export section at the bottom of report pages, gene lists, sequence pages, and method results pages has been improved and is now clearly arranged.

Figure 1.15. Export from a report page

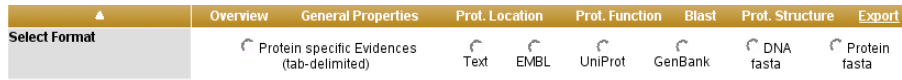


Figure 1.16. Export from a gene list



Figure 1.17. Export from a sequence page

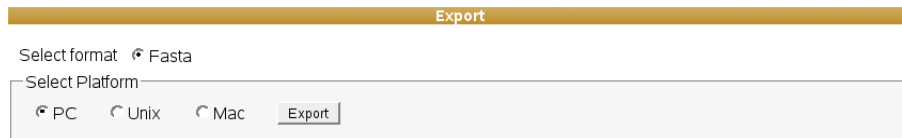


Figure 1.18. Export from a method results page



Improved Export formats

Pedant-Pro supports the export of data in a number of formats. For all formats the specifications for data export have been improved. Data exports in the following well defined formats are now supported:

- EMBL
- Genbank
- UniProt
- Fasta
- Text
- NCBI blast XML format

Improved default names for exported files

Per default exported files are now named as described in the table below:

Table 1.1. Naming of exported files

File format	File Extension	Example
Embl	.embl	gene1_contig1.embl

File format	File Extension	Example
Genbank	.genbank	gene1_contig1.genbank
UniProt	.uniprot	gene1_contig1.uniprot
Fasta	.fa	gene1_contig1.fa
Text	.txt	gene1_contig1.txt

Improved Command Line Interface data export

Export of analysis data for a complete analysis (e.g. a complete genome) in defined formats is supported by the Pedant-Pro Command Line Interface (CLI). Export is supported for the following formats:

- EMBL
- Genbank
- UniProt
- Fasta
- XML

Apollo Viewer is not longer supported

The Apollo Genome Annotation Curation Tool for viewing and modifying gene models is not longer supported.

New main service features

The following features of the main Pedant-Pro service have been added or modified:

- The conceptual distinction between primary workflow and secondary workflows has been removed. New method nodes can now be added directly to "the" workflow by exactly specifying the node dependencies.
- The object relational mapping framework Hibernate has been updated to version 3.2. The whole database abstraction layer has been adopted to support this major release upgrade.
- An additional adaptor to utilize compute grids has been implemented. The DRMAA (Distributed Resource Management Application API) adaptor allows to communicate with DRMAA conforming compute grids, e.g. Sun Grid Engine.

Chapter 2. Contact Biomax

For more information about the Pedant-Pro suite and the changes implemented in this version, contact Biomax:

Biomax Informatics AG
D-82152 Planegg
Robert-Koch-Str. 2
Germany
Tel.: +49 89 895574-0
Fax: +49 89 895574-825
Email: <info@biomax.com>
Website: www.biomax.com/pedantpro [<http://www.biomax.com/pedantpro>]

Biomax, Pedant-Pro, and BioRS are registered trademarks of Biomax Informatics AG in Germany and other countries. Java, Solaris and Sun are trademarks of Sun Microsystems, Inc. Microsoft and Windows are registered trademarks of Microsoft Corporation. Linux is a registered trademark of Linus Torvalds. SUSE is a registered trademark of Novell, Inc. Intel is a registered trademark and Xeon is a trademark of Intel Corporation. AMD and AMD Opteron are trademarks of Advanced Micro Devices, Inc. UNIX is a registered trademark of The Open Group. Mozilla and Firefox are registered trademarks of the Mozilla Foundation. Red Hat is a registered trademark of Red Hat, Inc. Debian is a registered trademark of Software in the Public Interest, Inc. Registered names, trademarks, etc., used in this document, even when not specifically marked as such, are not to be considered unprotected by law.