

## Selected genome annotation projects performed with the Pedant-Pro™ Sequence Analysis Suite

**Pedant-Pro Sequence Analysis Suite** The Pedant-Pro™ Sequence Analysis Suite helps research and development (R&D) departments save time and money by quickly extracting detailed and reliable information from sequences and presenting the data to scientists with a highly intuitive graphical interface.

The Pedant-Pro suite performs high-quality, comprehensive, enterprise-scale gene and genome analysis. The collected genetic information represents a central sequence, target and protein repository. For each gene and protein, a comprehensive report provides an overview of relevant features with links to public and proprietary databases. The report can be used to integrate diverse types of corporate data (e.g., expression analysis data).

The screenshot displays the Pedant-Pro Sequence Analysis Suite interface. The left sidebar shows a tree view of the report's contents, including Sequences, Protein function, BLAST results, Alignments, Protein structure, and Protein location. The main window shows a detailed report for the gene **gij27904513|ref|NC\_004545.1|.gene5**, identified as **S. ATP synthase chain (EC 3.6.3.14) (ATPase protein 6)**.

Overview		Protein function	Blast	Protein structure	Protein location	General properties	Export	
<b>Contig name</b>	gij27904513 ref NC_004545.1							
<b>Gene position</b>	2688-3488							
<b>ORF prediction method</b>	Gene prediction (Glimmer)							
Automatic FunCat (blastp FunCat)		Biological Processes						
	02.11	electron transport and membrane-associated energy conservation					1.4E-74	
	02.13	respiration					1.4E-74	
	02.45.15	energy generation (e.g. ATP Synthase)					1.4E-74	
	13.01.01.03	homeostasis of protons					1.4E-74	
Localization		40.30 prokaryotic cell membrane (inner membrane of gram - bacteria)						1.4E-74
Function		67.04.01.02 other cation transporters (Na, K, Ca, NH4, etc.)						1.4E-74
	67.15	electron / hydrogen carrier					1.4E-74	
	67.50.22	transport ATPases					1.4E-74	
Metacat obtained from EC		metabolism						
	01.03.01.27	ATP generation from ADP (oxidative phosphorylation)					3E	
	01.04.01.01	oxidative phosphorylation, photosynthesis					3E	
energy		02.11.01.01 ATP synthesis (proton transport)						3E
EC numbers		3.6.1.34 UNIPROT:Q8Z9S1 YERPE now EC 3.6.3.14						3.6E-80
		3.6.3.14 UNIPROT:ATP6_BUCBP H+-transporting two-sector ATPase 2.7E-109						
GO		Molecular function						
		GO:0003674	molecular_function					
		GO:0003824	catalytic activity					
		GO:0005215	transporter activity					
		GO:0005396	carrier activity					
		GO:0008324	cation transporter activity					

The Pedant-Pro gene report provides a comprehensive overview of all features of a gene

Biomax currently offers over 470 genomes annotated by the Pedant-Pro Sequence Analysis Suite, the PEDANT™ Genome Database. Numerous life science companies and research organizations regard the genomes as valuable repositories for gene and protein information. For details visit the Biomax website, <http://www.biomax.com/pedantgenomes>.

For information about the PEDANT genomes available from the Institute for Bioinformatics (IBI, the former Munich Information Center for Protein Sequences, MIPS) visit the website, <http://pedant.gsf.de>. For example, see the *Escherichia coli* K12 genome presented as Pedant-Pro 3 analysis, [http://pedant.gsf.de:3345/pedant3htmlview/pedant3view?Method=start\\_method&Db=Escherichia\\_coli\\_K12](http://pedant.gsf.de:3345/pedant3htmlview/pedant3view?Method=start_method&Db=Escherichia_coli_K12)

The PEDANT database system is reviewed in the *Nucleic Acids Research* database issue.

**References:**

**Riley ML et al (2007) PEDANT genome database: 10 years online. *Nucleic Acids Res* 35:D354–7**

**Riley ML, Schmidt T, Wagner C, Mewes HW and Frishman D (2005) The PEDANT genome database in 2005. *Nucleic Acids Res* 33:D308–10**

The Pedant-Pro suite has been used in a variety of genome annotation projects in academic and commercial environments. The following sections provide a sample overview of these projects.

**Bacteria** Numerous projects for the annotation of bacterial genes and genomes have been performed.

- The genome of an extra-intestinal pathogenic *Escherichia coli* strain was automatically annotated using the Pedant-Pro suite and other tools. Subsequently it was manually annotated by a team of Biomax experts. The project revealed the genetic basis of the urovirulence of this strain by showing the genetic composition (e.g., pathogenicity islands).

**Reference:**

**Brzuszkiewicz E et al (2006) How to become a uropathogen: comparative genomic analysis of extra-intestinal pathogenic *Escherichia coli* strains. *Proc Natl Acad Sci USA* 103:12879–84**

- The genome of the anammox (anaerobic ammonium oxidation) bacterium *Kuenenia stuttgartiensis* was annotated using the Pedant-Pro suite to discover the organism's special properties. Most significantly, candidate genes responsible for ladderane biosynthesis and biological hydrazine metabolism were identified and unexpected metabolic versatility was discovered.

**Reference:**

**Strous M et al (2006) Deciphering the evolution and metabolism of an anammox bacterium from a community genome. *Nature* 440:790–4**

- Genome analysis of a chlamydia-related symbiont of free-living amoebae was performed using the Pedant-Pro suite. It was shown that the last common ancestor of pathogenic and symbiotic chlamydiae (about 700 million years ago) was already adapted to intracellular survival in early eukaryotes and contained many virulence factors found in modern pathogenic chlamydiae. Chlamydiae are the major cause of preventable blindness and sexually transmitted disease.

**Reference:**

**Horn M et al (2004) Illuminating the evolutionary history of Chlamydiae. *Science* 304:728–30**



<http://www.biomax.com/pedantpro/>

- Pedant-Pro annotation of the genome of *Desulfotalea psychrophila*, a sulfate-reducing proteobacterium from permanently cold marine sediments, was used to reveal metabolic properties of this bacterial species. The bacteria are known to provide an essential contribution to the global carbon and sulfur cycles, and the analysis could show the genetic basis for characteristic pathways.

**Reference:**

**Rabus R et al (2004) The genome of *Desulfotalea psychrophila*, a sulfate-reducing bacterium from permanently cold Arctic sediments. *Environ Microbiol* 6:887**

- The Pedant-Pro Sequence Analysis Suite was used for the complete automatic annotation of *Pirellula sp.*, an aerobic, heterotrophic, marine species of the globally distributed and environmentally important bacterial order Planctomycetales.

**Reference:**

**Glöckner FO et al (2003) Complete genome sequence of the marine planctomycete *Pirellula sp.* strain 1. *Proc Natl Acad Sci USA* 100:8298–303**

- The Pedant-Pro Sequence Analysis Suite was used to identify lipoprotein genes in gram-positive bacteria.

**Reference:**

**Sutcliffe IC and Harrington DJ (2002) Pattern searches for the identification of putative lipoprotein genes in gram-positive bacterial genomes. *Microbiology* 148:2065–77**

The genomes of the following bacteria were manually annotated based on initial Pedant-Pro analysis:

- *Bacillus clausii* (2000)
- *Bacillus subtilis* (2002)
- *Chlamydia pneumoniae* (1999)
- *Corynebacterium glutamicum* (2000)
- *Escherichia coli* (2003)
- *Helicobacter pylori* J99 (July 2002)
- *Helicobacter pylori* KE26695 (ATCC 700392) (July 2002)
- *Listeria innocua* Clip11262 (May 2002)
- *Listeria monocytogenes* EGD (March 2002)
- *Mycobacterium tuberculosis* CDC 155 (in preparation)
- *Propionibacterium freudenreichii* (2001)

The Pedant-Pro software was used from 2000 to 2003 for the annotation of pathogenic bacteria in a network of 15 academic groups in Germany.



**Fungi** The following are the most important projects for annotation of fungal genomes with the Pedant-Pro suite.

- The genome of the crop-pathogenic fungus *Fusarium graminearum*, sequenced by the Whitehead Center for Genome Research, was automatically annotated and manually refined using the Pedant-Pro suite. The results are publicly available in the MIPS *Fusarium graminearum* Genome Database (FGDB, see URL below).

**Reference:**

<http://mips.gsf.de/genre/proj/fusarium/>

The *Fusarium graminearum* genome was analyzed and annotation performed using the Pedant-Pro suite. Some interesting features that were revealed are the following: the genome shows very few repetitive sequences, low repeat content, and low number of paralog (ancestrally duplicated) genes. For another analyzed strain of *F. graminearum*, more than 10,000 single nucleotide polymorphisms (SNPs) were detected. A great number of highly polymorphic regions contained gene sets which are presumably related to plant-fungus interactions.

**Reference:**

**Cuomo CA et al (2007) The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science* 317:1400–2**

The annotated *Fusarium graminearum* genome also provided the basis for the design of an Affymetrix chip.

**Reference:**

**Gldener U et al (2006) Development of a *Fusarium graminearum* Affymetrix GeneChip for profiling fungal gene expression in vitro and in planta. *Fungal Genet Biol* 43:316–25**

- The Pedant-Pro suite was used in genome analysis of the filamentous fungus *Aspergillus niger*, which is widely used in the food industry.

**Reference:**

**Pel HJ et al (2007) Genome sequencing and analysis of the versatile cell factory *Aspergillus niger* CBS 513.88. *Nat Biotechnol* 25:221–31**

- The Comprehensive Yeast Genome Database (CYGD), which includes the genomes of several yeast species analyzed with the Pedant-Pro system, is published. The database provides information about the cellular functions of *Saccharomyces cerevisiae* and related species, which are typical eukaryotic model organisms and references for new annotations in the taxon range.

**References:**

**Güldener U et al (2005) CYGD: the Comprehensive Yeast Genome Database. *Nucleic Acids Res* 33: D364–8**

<http://mips.gsf.de/genre/proj/yeast/>



<http://www.biomax.com/pedantpro/>

- The Pedant-Pro Sequence Analysis Suite was used for extensive protein analysis for all predicted genes of the *Neurospora crassa* genome. The project provided an overview of the protein equipment of a filamentous fungus and showed that *Neurospora crassa* produces proteins which are typical for filamentous fungi and other proteins which are typical for prokaryotes. The project formed the basis for the MIPS *Neurospora crassa* database (MNCDB).

**References:**

**Galagan JE et al (2003) The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* 422:859–68**

**Mannhaupt G et al (2003) What's in the genome of a filamentous fungus? Analysis of the *Neurospora* genome sequence. *Nucleic Acids Res* 31:1944–54**

<http://mips.gsf.de/projects/fungi/neurospora>

- The Pedant-Pro Sequence Analysis Suite was used for automatic gene extraction during the complete genome characterization of *Hansenula polymorpha* (strain CBS4732).

**Reference:**

**Ramezani-Rad M et al (2003) The *Hansenula polymorpha* (strain CBS4732) Genome Sequencing and Analysis. *FEMS Yeast Research* 4:207–15**

The genomes of the following fungi were manually annotated based on initial Pedant-Pro analysis.

- *Aspergillus niger* (2001–2002)
- *Hansenula polymorpha* (2003)
- *Penicillium chrysogenum* (2004–2005)

**Human** In 2001 Biomax released the first version of the Biomax Human Genome Database (BHGDB) Web portal, which contains manual annotation of the complete human genome. The automatic analysis used as the basis for the manual annotation was performed using the Pedant-Pro Sequence Analysis Suite. An updated version of the BHGDB will be available soon.



**Reference:**

**Schüller C and Fritz A (2002) An enhanced Human Genome Database. *Genet Eng News* 22:38**

**Plants** The following *Arabidopsis thaliana*-related projects were performed using the Pedant-Pro suite.

<http://www.biomax.com/pedantpro/>

- The first version of the MIPS *Arabidopsis thaliana* database (MAtdB), which includes Pedant-Pro-based annotation, is published. MatDB serves as an information pool containing diverse functional genomic data about this important model plant.

**References:**

**Schoof H et al (2004) MIPS *Arabidopsis thaliana* Database (MAtdB): an integrated biological knowledge resource for plant genomics. *Nucleic Acids Res.* 32:D373–6**

**Schoof H et al (2002) MIPS *Arabidopsis thaliana* Database (MAtdB): an integrated biological knowledge resource based on the first complete plant genome. *Nucleic Acids Res* 30:91–3**

<http://mips.gsf.de/proj/thal/db>

- The complete genome of *Arabidopsis thaliana* was published in Nature.

The *A. thaliana* genome is available as a PEDANT database on compact disk (CD) with the Pedant-Pro graphical user interface.

**Reference:**

**Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796–815**



- Pedant-Pro analysis of *Arabidopsis thaliana* Chromosome 4 is completed.

**Reference:**

**The European Union Sequencing Consortium and The Cold Spring Harbor, Washington University in St. Louis and PE Biosystems Arabidopsis Sequencing Consortium (1999) Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. *Nature* 402:769–77**

**Animals** The following are the most important Pedant-Pro analyses of mammalian genomes.

- The mouse (*Mus musculus*) genome was manually annotated based on an initial Pedant-Pro analysis and is published as the MIPS Mouse Functional Genome Database.

**References:**

**Ruepp A et al (2006) The Mouse Functional Genome Database (MfunGD): functional annotation of proteins in the light of their cellular context. *Nucleic Acids Res* 34:D568–71**

<http://mips.gsf.de/genre/proj/mfungd/>

- The Pedant-Pro suite is used for the Cow EST Project and the International Pig Genome Project at the Korean National Livestock Research Institute (NLRI).

**Reference:**

**Biomax press release (2005),**

[http://www.biomax.com/company/pressrel/2005\\_nlri.pdf](http://www.biomax.com/company/pressrel/2005_nlri.pdf)

**Literature overview** The following publications marked the first two versions of the Pedant-Pro application. A **Pedant-Pro software** scientific paper about version 3 will be prepared soon.

- The Pedant-Pro Sequence Analysis Suite version 2 is released. Version 2 of the software includes the upstream gene predictions in addition to protein analysis. The entire pipeline from raw sequence data to potential protein function is covered, which can save scientists a lot of time and effort.

**Reference:**

**Frishman D et al (2001) Functional and structural genomics using PEDANT. *Bioinformatics* 7:44–57**

- The first version of the Protein Extraction Description and Analysis Tool (PEDANT) system, which provides automatic annotation of proteins using numerous state-of-the-art bioinformatics algorithms, is published. The system is the first to incorporate a wide spectrum of methods with an easy-to-use graphical interface.

**Reference:**

**Frishman D and Mewes HW (1997) PEDANTic genome analysis. *Trends Genet* 13:415–6**

Biomax and Pedant-Pro are registered trademarks of Biomax Informatics AG in Germany and other countries. Registered names, trademarks, etc., used in this document, even when not specifically marked as such, are not to be considered unprotected by law. PDPROSUST0709

