#### Editing Pathway/Genome Databases II

#### By Ron Caspi

ron.caspi@sri.com



A copy of this presentation could be found at http://bioinformatics.ai.sri.com/ptools/tutorial/sessions/curation

A lot more information is available in the Curator's Guide, at <u>https://bioinformatics.ai.sri.com/ptools/curatorsguide.pdf</u>

#### Summary of last tutorial

What we discussed:

- Starting the software from the Lisp window
- Switching between the main window and Lisp window, breaks
- Data structure: frames, classes and instances
- Some menu commands
- Author credit system, creating curator and organization frames
- The compound, reaction, protein, and pathway editors
- Avoiding duplication

#### Summary of this tutorial

What we will discuss today:

- Gene editor
- Creating protein complexes
- Writing summaries, using citations, internal hyperlinks
- Editing transcription units
- Curating regulatory interactions
- PGDB refinement operations
- Propagating MetaCyc updates
- The Consistency Checker

Once again, we will be using the PGDB for *Arthrospira platensis NIES-39* for practicing



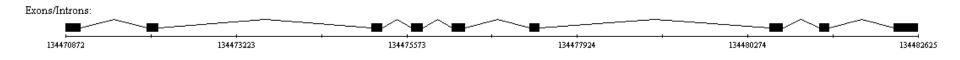
#### The Gene and Isoform/Coding-Segment Editors

- Enter synonyms/accession numbers
- Enter Links to other databases
- Define transcription direction
- Modify start/end positions
- Define introns
- Create new isoforms
- Define frame shifts

Isoform/Coding-Segment Editor
Isoform Gene Product:
Base numbers are: Re arsenate reductase
Create New Isoform
If specifying multiple coding segments, please provide the appropriate interpret
RNA splicing: The RNA transcript is spliced to remove introns.
Protein splicing: The immature pro-peptide is spliced to remove inteins.
Ribosomal slippage: The ribosome slips during translation to generate a pro-
You do not need to supply an interpretation here if you are merely using this fo
Interpretation: Not specified 🔹
,

Class: Genes						
ommon Name: asd	2		_			
nonyma						
asd2			Make this the Com	mon Name	í I	
ccession-1. GBAA	3937					
ccession-2						
roduct aspartate-ser	saldehude dehudran	1000 Edg				
tromosome Chromo	1100 C 100 C 100 C 100 C					
inks to other database	Contraction of the local sectors of the local secto					
	atabase	D	Relationship	2		
NCEI Entrez-Gene		<ul> <li>AJH87363</li> </ul>	Same Entity			
		•	Same Entity			
Citations:	_				-	
	00.001/07.11	entified by sequence				_
ranscription Direction	- (reverse) 💌	Start Position: 3614	777 End	Position: 3	613731	
equence ATOGAAJ CAAATOT	аяс алалаастті таа агастітая	T TCATGTOGOT GTAG	TTGGAG CAACOSS CAATOS GGAAGTI	DODC AGTT	OGTGAA	
едиелсе: Атобаал САААТОТ ТСТАААС	аос алалаластті тал атастітася зат стослоотал	TCATGTOGCT GTAG	TTGGAG CAACOSS CAATOS GEAAGTT AAGOCS AAGAATT	DODC AGTI TAAC GTIX	CTITCA CTITCA	
equence ATGGAAJ CAAATGT TCTAAAC GCAACTG TCAAAAC	AGC ARAAAACTTI TAA AIACTIIAGA MAT CIGCAGGIAA CIG AAAGTIIIGA AAT IAGOSCCAGA	T TCATGTOGOT GTAG GAAACGAGAA TITO GAAACTTGTA TITA AGGAGTAGAT ATOS ASCAGCAAAG COOS	TTGGAG CAACOD CAATOG GGAAGTI ANGGOG AAGAATI CACTAT TINGTGC GIGOGA TIGITGI	PODC AGTT TAAC GTT2 TTAC AGTT CAGG TOGJ TTGA TAAT	COSTGAA CTITICA CAAGAG LICTIGTA RCAAGT	
equence: ATOGAAJ CAAATOT TCTAAAC GCAATO TCAAAAC GCATOC AAAGAAC	АОС АЛАЛАААСТТІ ТАА АІАСТІІАОР КАТ СТОСАЮБІАЛ СТО АЛАБІІІТОЛ ААТ ТАЙСЬССАОР КІТА ТОАСАОЛАЛІ АТА АТОБІАТТАІ	TCAIGTOSCT STAG SAAACGAGAA TITC GAAACTIGIA TITC AGGAGTAGAT ATCS ASCAGCAAR ACCS COTOCCACTT STIC TGCAMATCCA AACT	TTOGAG CAACOSS CAATOG GGAAGTI AAGGCG AAGAATI CACTAT TIAGTGC GTGOGA TIGTTGI TACCTG AAGTAJJ GTTCTA CAATTGI	SOSC AGTI IAAC GIII IIAC AGTI IIAC AGTI IIGA IAAI AIGA AAAC AAAT GGI	POGTIGAA ACTITICA PCAAGAG AICTOTA PACAAGT PACAAGT PACTTA AGTAOCT	
едиелсе: Атодлал Смалтот тоталал боласти толалал боласти алалала сттолог	AGC AAAAAACTTT TAA ATACTTTAGA MAT CTGCAGGTAA CTG AAAGTTTTAG MAT TAGCGCCAGA MITA TGACBGAAAA ATA ATGGTATAT CAG TCGTCAGCA	TCATGTOGET GIAG GAAACIAGAA TITC GAAACITGIA TITA AGGAGIAGAT ATCG AGGAGIAGAA COOS A OFIDECAATS GIGS TGCAAAGICA AAC AIAIGGIITA AAAC	ТТОБАС СААСООГ СААТОВ ОБАЛЯТІ АЛОВОВ АЛОЛАТІ САСТАТ ТІЛІТОС ОТОСАТ ТІЛІТОС ТАССТО АЛОТАЛ ОТТСТА СААТІСЯ БАЛТАЛ ТОБТІТС	SOSC AGTI IAAC GIIX IIAC AGTI CASS TOSS IIGA IAAI ATGA AAAC AAAT GGIX DCAC ATAI	OGTIGAA ICTITICA ICAAGAG ILCTOTA IACAAGT IGAACTTA IGTAOCT ICAAGCT	
equence ATGUAAT CAAATGT TCTAAAA GCAACTG TCAAAAA GCATTGI AAAAAAA GTTGAGG GTATCAA AATGGG	AGC AAAAAACTTT TAA ATACTTTAGA GAT CTGCAGGTAA AT TAGCGCCAGA GTA TGACRGAAAA ATA ATGGTACTAGT CAG TCGTCAGG AGG CTGGTGCGGG AAG AAGTGAAGGC	Г ТСАТОТОВСТ БТАБ БЛААССЛОГАТ ТТТА 4 ЛОДАСТАТАТА ТТТА 4 ЛОДАСТАГАТ ТТТА 4 ЛОДАСТАГАТ АТСО 5 ОСТОВСЕКТ БТТО 7 ТОСАМАТСКА АМСТ 6 ЛОСОВАТСАЯ БАМС 6 ЛОСОВАТСЯЯ БАМС 6 ЛОСОВАТСЯЯ СЛО	TTGGAG CAACOSE CAATOG GGAAGTI ANGGOG AACAATI CACTAT TIAGTOG TGGGA TIGTTGI TACCTG AACTAI GITCTA CAATICS GAGTAA TCGITCA TICATG AACTAI TATCAG GTGAIAS	SOSC AGTI IAAC GTI2 ITAC AGTI CASS TOSS ITGA TAAI ATGA AAA CAAAT GUT2 COAC ATAI CACA AGCI AGAA ACAI	IGGTGAA ICTITCA CCAADAG ITCTGTA ACAAGT IGACTTA ICTAOCT ICAAGCT IAITCTTA ITTTOCA	
equence ATGUAAI CAAAI COAAI CTAAAA GCAAI CTAAAA CIIGAS GTAICA AAIGOO AAIGOO ATGOCT	АКС АЛАЛААСТТІ ТАЛ АІЛСТІТТАЙ ЧАТ СТОСЛЮТЛА СТО АЛАГІТТІЙ ЗАЛ ТАЮОВОСЛАВ ЗАЛ АТООГАТТАЇ СКА ТТООГАТТАЇ СКА ТТООГАТТАЇ СА ТТООГАЛІОС АЛА АЛОГОЛАВО АЛАТОЛАВО	I TCATGTOGUT GTAG GAAACTAGTA TTTX AAGAGTAGTAT TTTA AGGAGTAGTAT ATON AGGAGTAGTAGTA GUTGOCACTT GTTG TGCAATCCA AACT ATATGGTTTA AAAC AGGATTGAA GAAC	ТТОБАБ СААСОВІ СРАТОФ ФІРАДТІ САСТАТ ТІЛЬТОС ОТФОДА ІТОТТО ТАСТІС АЛЬТАЗ БІЛТСТА СААТІСЯ БІЛТСТА СААТІСЯ БІЛТСТА СААТІСЯ ТІТСАТА ФІСАЛІТІ ТІСАТА ФІРАЛАТОГІ	SOSC AGTI TAAC GTI2 TAAC AGTI CASS TOSS TIGA TAAI ATGA AAAT GCAC ATAI CACA AGCI AGAA ACAI BATT TAC	IGOTGAA ICTITCA ICAAGAG ICTUTA ACCAGT IGACTTA IGTAGCT ICAAGCT ITTICCA ITTICCA	
equence:	АОС АЛАЛАСТТІ ТАА АТАСТІТАФ АІТ СТОСАЮТА СТО АЛАСТІТТОЙ АЛАТ ТАССОССАВ ТА ТАССОССАВА ТА ТОСОТАФАТ ТА ТОСОТАФАТ ТОСОТАФАТ АЛАТ АЛОГАЛОВ ТСА МОВСАТТОС ДАЛА ТОЛІТАЛОВ ТО ТАСОТТІЛО	I TCAIGTOSCI GIAG GAAACGAGAA TITO GAAACTOTA TITA AGGAGGAGAT ATO A DEGAGGAGAT ATO TOCAAATOCA AACT AGCAGGAATGAA GAAC GAACGATTA AAACT ACCAGATGAA AACT AACGAATGAA ATA	TTODAG CALCOSE CAATCG GGAGTT ANGOOG AAGAAT ANGOOG AAGAAT CACTAT TTAGTOC CTOGGA TTGTTG CAGTAA TOGTTG TACTGA ACCAAT TCCATG AACAATG TCCATG AGTCGT ACTCGG AGTCGT	SOSC AGTI TAAC GIII CAGG TOG TIGA TAG TIGA TAG ALTGA AGC ALTA CACA AGC AGAT TAC ALTI AGAT	NOTGAA CITITCA YAAGAG ITCTGTA ACAAGI MACTITA IGTAOCT YATCITA TITTCA HITTGAA IGTAOCG GAAGITA	
единсе Атобал Салатот ТСТАЛА ССАЛТС ССАЛТС Аласал СТТСАЛ СТТСАЛ СТТСАЛ СТТСАЛ СТТСАЛ СТТСАЛ СТТСАЛ СТТСАЛ СТАТСС АЛТОСТ САЛТСТ САЛТСТ САЛТСТ САЛТСТ САЛТСТ ССАЛТС ССАЛТС ССАЛТС ССАЛТС ССАЛТСТ СССАЛТС СССАЛТС ССАЛТС СССС ССС	AGC AMAMAACTIT TAA ATACTITAGA AT CIGCAGOTAA CIG AAAGITTIGA AT TACCOCCAS SITA TEACAGAAAA ATA ATCOTATATA CAG TICOTCACCA SITG CIGGISCOSC AAG AAATCAAGO TAA TOATTAACG SITG TACGITTACC SITG TACGITTACCA SITG CAGAACAET	I TCAIGTOSCI GIAG GRAACGMGAA TITIC OLAACTIGTA TITA AGGAGTAGAA ATCA AGCAGTAGAA COOS OCTOOCACHT GIIG TOCAMATICA AACT ACAGATITA CAG CACAGITITA CAG ACCAGATIGAT AAGT ACGAATAGA AATA	TTODAG CARCOUS CARLOS GUAGTI AGOST AGADATI CACTAT TTATOS TODOA TEGTTO TOCTO AGTAN GUIDA CATTOS AGUTA CONTO TECATO AGADTO TECATO TOCTOS ACTOTO AGUTOS ACTOTO AGUTOS	SCOC AGTI TAAC GTIJ TTAC AGTI TTAC AGTI TTAC AGTI AGTA AAAC AGTA AGTA AGTA AGTA AGTA AGTA AGTA	COTGAA CTITCA CAAGAG ICTGTA ACCAGT SGACTTA GTAGCT CAAGCT ATCTTA ITTTCA ITTTCA ITTTCA ITTTCA ISTAGCG GAAGTA GGTATC	e.
единсе СААЛТЯ ССААЛТЯ ТСТАХАА ССААСТ ТСТАХА СААТТО СААЛТО СТТОКО СТТОКО СТТОКО СТТОКО СТТОКО СААСТА СААСТА СААСТА	АОС АКААААСТІТ ТАА ХІАСТІАВ ОБАТ СІĞСЬІБТАЯ КТІ АЛАСТІТІАВ АЛТ ТАРСОССАВ ОТА ТЕЛСІВЛАНА ТАРСОССАВ ОТА ТЕЛСІВЛАНА ТАЛТАКА ТІСКІВОСАТІТАС АЛВ АЛАТТАЛА АЛАТТАЛА ОСТАЛСТАВ АЛАТТСОСОВАЯ	T CANOTCOUT OTAG T CANACIDAT TITO CANACIDATA TITO CANACIDATA TO A SCINCTURA TATO A SCINCTURA TO COTOCINITY AND COTOCINITY AND CONSTITUTA CONT CANOTITIA CONT CANOTITIA CONT ACOUNTINA CON CANOTITIA CON CONTOCINITY ACOUNTINA CON CONSTITUTA AND ACOUNTINA AND	ТТОВАВ САЛСОВО САЛТОВ ФИЛАНТІ АЛБОСИ АЛБАЛТІ ТЛОТОВ АЛБАЛТІ ТЛОТОВ АЛБАЛТІ ТЛОТОВ АЛБАЛТІ ТЛОТОВ АЛБАЛТІ ТОСАЛЯ АТАЛТОВ ТОСАЛЯ АТАЛТОВ СОЛТАС ТТОСЛАЯ САЛТОВ САЛСТАЯ САЛТОВ САЛСТАЯ	SCSC AGTT TAAC GITA TTAC AGTT TTAC AGTT TTAC AGTT TTGA TAAAT TCGA AGTA AAT GGTA AAT GGTA GATT TACC AATT AGAA TTTA CATT ATTC COAA TTCA AGTA AASG ATTC	INTERNA INTERNA INTERNA INTERNA INCLASSI INTERNA INTER	¢.
яриное СААЛТО ССААЛТО ТСТАААА ОСАЛТО ТСТАААА СААЛТО СТТОВА СТТОВА СТТОВА СТТОВА ОДАЛТО ССААСТ ОДАЛТО ОДАЛТО ОДАЛТО ОДАЛТО ОДАЛТО ОДАЛТО ОДТОТА ХАСТВАЙ	АОС АКААААСТІТ ТАА ХІАСТІАВ ОБАТ СІĞСЬІБТАЯ КТІ АЛАСТІТІАВ АЛТ ТАРСОССАВ ОТА ТЕЛСІВЛАНА ТАРСОССАВ ОТА ТЕЛСІВЛАНА ТАЛТАКА ТІСКІВОСАТІТАС АЛВ АЛАТТАЛА АЛАТТАЛА ОСТАЛСТАВ АЛАТТСОСОВАЯ	TCAIDTOOCT GIAG GAAACHAAA TITO GAAACHAAA TITO GAAACHAAA TITO GAAACHAAA GOTMICLAIT ATO GOTMICLAIT GITO TOCAAACHAA ACTUGITTI AAAC ACCAANTINA TAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA AAAAAATA AAA GAACHATTA AAAA AAAAAAATA AAAA AAAAAATA AAAAAAAA	ТТОВАВ САЛСОВО САЛТОВ ФИЛАНТІ АЛБОСИ АЛБАЛТІ ТЛОТОВ АЛБАЛТІ ТЛОТОВ АЛБАЛТІ ТЛОТОВ АЛБАЛТІ ТЛОТОВ АЛБАЛТІ ТОСАЛЯ АТАЛТОВ ТОСАЛЯ АТАЛТОВ СОЛТАС ТТОСЛАЯ САЛТОВ САЛСТАЯ САЛТОВ САЛСТАЯ	SCSC AGTT TAAC GITA TTAC AGTT TTAC AGTT TTAC AGTT TTGA TAAAT TCGA AGTA AAT GGTA AAT GGTA GATT TACC AATT AGAA TTTA CATT ATTC COAA TTCA AGTA AASG ATTC	INTERNA INTERNA INTERNA INTERNA INCLASSI INTERNA INTER	e.
яриное СААЛТО ССААЛТО ТСТАААА ОСАЛТО ТСТАААА СААЛТО СТТОВА СТТОВА СТТОВА СТТОВА ОДАЛТО ССААСТ ОДАЛТО ОДАЛТО ОДАЛТО ОДАЛТО ОДАЛТО ОДАЛТО ОДТОТА ХАСТВАЙ	АОС АХААААСТІТ ТАА АТАСТІТАВ ОБАТ СТОСАЮТАЯ СТО АЛАГІТТІВ АЛ ТАЛОСССАВ ОТА ТОСКОССАВ ОТА ТОСКОССАВ ОТА АЛОСССАВ ОТА АЛОССАТА СТОСТОВО ОТО АЛОССАТА СТОСТОВО ОТО АЛОСТАТАТ ОТО АЛАГТАТТВ ОТО АЛАГТАТВ ОТО АЛАГТАТВ ОТО АЛАГТАТВ ОТО ОТО ОТО ОТО ОТО ОТО ОТО ОТО ОТО ОТО	TCAIDTOOCT GIAG GAAACHAAA TITO GAAACHAAA TITO GAAACHAAA TITO GAAACHAAA GOTMICLAIT ATO GOTMICLAIT GITO TOCAAACHAA ACTUGITTI AAAC ACCAANTINA TAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA AAAAAATA AAA GAACHATTA AAAA AAAAAAATA AAAA AAAAAATA AAAAAAAA	ТТОВАВ САЛСОВО САЛТОВ ФИЛАНТІ АЛБОСИ АЛБАЛТІ ТЛОТОВ АЛБАЛТІ ТЛОТОВ АЛБАЛТІ ТЛОТОВ АЛБАЛТІ ТЛОТОВ АЛБАЛТІ ТОСАЛЯ АТАЛТОВ ТОСАЛЯ АТАЛТОВ СОЛТАС ТТОСЛАЯ САЛТОВ САЛСТАЯ САЛТОВ САЛСТАЯ	SCSC AGTT TAAC GITA TTAC AGTT TTAC AGTT TTAC AGTT TTGA TAAAT TCGA AGTA AAT GGTA AAT GGTA GATT TACC AATT AGAA TTTA CATT ATTC COAA TTCA AGTA AASG ATTC	INTERNA INTERNA INTERNA INTERNA INCLASSI INTERNA INTER	E.
ADALAS ATGENAL CARATO TCTANA GCANTO TCTANA GCANTO CTTGRA GTATOR GANCIA GTATOR GANCIA GTTCTM TOPOTO	АОС АХААААСТІТ ТАА АТАСТІТАВ ОБАТ СТОСАЮТАЯ СТО АЛАГІТТІВ АЛ ТАЛОСССАВ ОТА ТОСКОССАВ ОТА ТОСКОССАВ ОТА АЛОСССАВ ОТА АЛОССАТА СТОСТОВО ОТО АЛОССАТА СТОСТОВО ОТО АЛОСТАТАТ ОТО АЛАГТАТТВ ОТО АЛАГТАТВ ОТО АЛАГТАТВ ОТО АЛАГТАТВ ОТО ОТО ОТО ОТО ОТО ОТО ОТО ОТО ОТО ОТО	TCAIDTOOCT GIAG GAAACHAAA TITO GAAACHAAA TITO GAAACHAAA TITO GAAACHAAA GOTMICLAIT ATO GOTMICLAIT GITO TOCAAACHAA ACTUGITTI AAAC ACCAANTINA TAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA AAAAAATA AAA GAACHATTA AAAA AAAAAAATA AAAA AAAAAATA AAAAAAAA	ТТОВАВ САЛСОВО САЛТОВ ФИЛАНТІ АЛБОСИ АЛБАЛТІ ТЛОТОВ АЛБАЛТІ ТЛОТОВ АЛБАЛТІ ТЛОТОВ АЛБАЛТІ ТЛОТОВ АЛБАЛТІ ТОСАЛЯ АТАЛТОВ ТОСАЛЯ АТАЛТОВ СОЛТАС ТТОСЛАЯ САЛТОВ САЛСТАЯ САЛТОВ САЛСТАЯ	SCSC AGTT TAAC GITA TTAC AGTT TTAC AGTT TTAC AGTT TTGA TAAAT TCGA AGTA AAT GGTA AAT GGTA GATT TACC AATT AGAA TTTA CATT ATTC COAA TTCA AGTA AASG ATTC	INTERNA INTERNA INTERNA INTERNA INCLASSI INTERNA INTER	E.
ADALAS ATGENAL CARATO TCTANA GCANTO TCTANA GCANTO CTTGRA GTATOR GANCIA GTATOR GANCIA GTTCTM TOPOTO	АОС АХААААСТІТ ТАА АТАСТІТАВ ОБАТ СТОСАЮТАЯ СТО АЛАГІТТІВ АЛ ТАЛОСССАВ ОТА ТОСКОССАВ ОТА ТОСКОССАВ ОТА АЛОСССАВ ОТА АЛОССАТА СТОСТОВО ОТО АЛОССАТА СТОСТОВО ОТО АЛОСТАТАТ ОТО АЛАГТАТТВ ОТО АЛАГТАТВ ОТО АЛАГТАТВ ОТО АЛАГТАТВ ОТО ОТО ОТО ОТО ОТО ОТО ОТО ОТО ОТО ОТО	TCAIDTOOCT GIAG GAAACHAAA TITO GAAACHAAA TITO GAAACHAAA TITO GAAACHAAA GOTMICLAIT ATO GOTMICLAIT GITO TOCAAACHAA ACTUGITTI AAAC ACCAANTINA TAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA AAAAAATA AAA GAACHATTA AAAA AAAAAAATA AAAA AAAAAATA AAAAAAAA	ТТОВАВ САЛСОВО САЛТОВ ФИЛАНТІ АЛБОСИ АЛБАЛТІ ТЛОТОВ АЛБАЛТІ ТЛОТОВ АЛБАЛТІ ТЛОТОВ АЛБАЛТІ ТОСЛАВ АГАЛТІ ТОСЛАВ АТАЛТОВ ТОСЛАВ АТАЛТОВ ОСТТАС ТРОГЛАВ САЛТОВ САВСТАВ	SCSC AGTT TAAC GITA TTAC AGTT TTAC AGTT TTAC AGTT TTGA TAAAT TCGA AGTA AAT GGTA AAT GGTA GATT TACC AATT AGAA TTTA CATT ATTC COAA TTCA AGTA AASG ATTC	INTERNA INTERNA INTERNA INTERNA INCLASSI INTERNA INTER	E.
ADALAS ATGENAL CARATO TCTANA GCANTO TCTANA GCANTO CTTGRA GTATOR GANCIA GTATOR GANCIA GTTCTM TOPOTO	АОС АХААААСТІТ ТАА АТАСТІТАВ ОБАТ СТОСАЮТАЯ СТО АЛАГІТТІВ АЛ ТАЛОСССАВ ОТА ТОСКОССАВ ОТА ТОСКОССАВ ОТА АЛОСССАВ ОТА АЛОССАТА СТОСТОВО ОТО АЛОССАТА СТОСТОВО ОТО АЛОСТАТАТ ОТО АЛАГТАТТВ ОТО АЛАГТАТВ ОТО АЛАГТАТВ ОТО АЛАГТАТВ ОТО ОТО ОТО ОТО ОТО ОТО ОТО ОТО ОТО ОТО	TCAIDTOOCT GIAG GAAACHAAA TITO GAAACHAAA TITO GAAACHAAA TITO GAAACHAAA GOTMICLAIT ATO GOTMICLAIT GITO TOCAAACHAA ACTUGITTI AAAC ACCAANTINA TAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA AAAAAATA AAA GAACHATTA AAAA AAAAAAATA AAAA AAAAAATA AAAAAAAA	ТТОВАВ САЛСОВО САЛТОВ ФИЛАНТІ АЛБОСИ АЛБАЛТІ ТЛОТОВ АЛБАЛТІ ТЛОТОВ АЛБАЛТІ ТЛОТОВ АЛБАЛТІ ТОСЛАВ АГАЛТІ ТОСЛАВ АТАЛТОВ ТОСЛАВ АТАЛТОВ ОСТТАС ТРОГЛАВ САЛТОВ САВСТАВ	SCSC AGTT TAAC GITA TTAC AGTT TTAC AGTT TTAC AGTT TTGA TAAAT TCGA AGTA AAT GGTA AAT GGTA GATT TACC AATT AGAA TTTA CATT ATTC COAA TTCA AGTA AASG ATTC	INTERNA INTERNA INTERNA INTERNA INCLASSI INTERNA INTER	8
equence CAAATO TOTAAA GCAATO TOTAAA GCATO TOTAA GAATO GTATOA GAATO GTATOA GAATO GTATOA GAATO GTATOA GAATO GTATOA GTATOA GTATOA GTATOA	АОС АХААААСТІТ ТАА АТАСТІТАВ ОБАТ СТОСАЮТАЯ СТО АЛАГІТТІВ АЛ ТАЛОСССАВ ОТА ТОСКОССАВ ОТА ТОСКОССАВ ОТА АЛОСССАВ ОТА АЛОССАТА СТОСТОВО ОТО АЛОССАТА СТОСТОВО ОТО АЛОСТАТАТ ОТО АЛАГТАТТВ ОТО АЛАГТАТВ ОТО АЛАГТАТВ ОТО АЛАГТАТВ ОТО ОТО ОТО ОТО ОТО ОТО ОТО ОТО ОТО ОТО	TCAIDTOOCT GIAG GAAACHAAA TITO GAAACHAAA TITO GAAACHAAA TITO GAAACHAAA GOTMICLAIT ATO GOTMICLAIT GITO TOCAAACHAA ACTUGITTI AAAC ACCAANTINA TAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA AAAAAATA AAA GAACHATTA AAAA AAAAAAATA AAAA GAACHATTA AAAA AAAAAATA AAAAAAAAAAAAAAAAAAAA	ТТОВАВ САЛСОВО САЛТОВ ФИЛАНТІ АЛБОСИ АЛБАЛТІ ТЛОТОВ АЛБАЛТІ ТЛОТОВ АЛБАЛТІ ТЛОТОВ АЛБАЛТІ ТОСЛАВ АГАЛТІ ТОСЛАВ АТАЛТОВ ТОСЛАВ АТАЛТОВ ОСТТАС ТРОГЛАВ САЛТОВ САВСТАВ	SCSC AGTT TAAC GITA TTAC AGTT TTAC AGTT TTAC AGTT TTGA TAAAT TCGA AGTA AAT GGTA AAT GGTA GATT TACC AATT AGAA TTTA CATT ATTC COAA TTCA AGTA AASG ATTC	INTERNA INTERNA INTERNA INTERNA INCLASSI INTERNA INTER	E.
GLANCE CAAATO CCAAATO CCAAATO TCTAAAA CCAAATO CCAACT CTCAAA AATOOC AATOOC CTTCAA AATOOC CTTCAA AATOOC CAACTA CAAATO CTTCAA CAAATO CTTCAA CTTCAA CTTCAA CTTCAA CTTCAA	АОС АХААААСТІТ ТАА АТАСТІТАВ ОБАТ СТОСАЮТАЯ СТО АЛАГІТТІВ АЛ ТАЛОСССАВ ОТА ТОСКОССАВ ОТА ТОСКОССАВ ОТА АЛОСССАВ ОТА АЛОССАТА СТОСТОВО ОТО АЛОССАТА СТОСТОВО ОТО АЛОСТАТАТ ОТО АЛАГТАТТВ ОТО АЛАГТАТВ ОТО АЛАГТАТВ ОТО АЛАГТАТВ ОТО ОТО ОТО ОТО ОТО ОТО ОТО ОТО ОТО ОТО	TCAIDTOOCT GIAG GAAACHAAA TITO GAAACHAAA TITO GAAACHAAA TITO GAAACHAAA GOTMICLAIT ATO GOTMICLAIT GITO TOCAAACHAA ACTUGITTI AAAC ACCAANTINA TAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA AAAAAATA AAA GAACHATTA AAAA AAAAAAATA AAAA GAACHATTA AAAA AAAAAATA AAAAAAAAAAAAAAAAAAAA	ТТОВАВ САЛСОВО САЛТОВ ФИЛАНТІ АЛБОСИ АЛБАЛТІ ТЛОТОВ АЛБАЛТІ ТЛОТОВ АЛБАЛТІ ТЛОТОВ АЛБАЛТІ ТОСЛАВ АГАЛТІ ТОСЛАВ АТАЛТОВ ТОСЛАВ АТАЛТОВ ОСТТАС ТРОГЛАВ САЛТОВ САВСТАВ	SCSC AGTT TAAC GITA TTAC AGTT TTAC AGTT TTAC AGTT TTGA TAAAT TCGA AGTA AAT GGTA AAT GGTA GATT TACC AATT AGAA TTTA CATT ATTC COAA TTCA AGTA AASG ATTC	INTERNA INTERNA INTERNA INTERNA INCLASSI INTERNA INTER	
204502 7 1750333 1757334 1757334 1757334 1757334 1757334 175744 17574 175744	АОС АХААААСТІТ ТАА АТАСТІТАВ ОБАТ СТОСАЮТАЯ СТО АЛАГІТТІВ АЛ ТАЛОСССАВ ОТА ТОСКОССАВ ОТА ТОСКОССАВ ОТА АЛОСССАВ ОТА АЛОССАТА СТОСТОВО ОТО АЛОССАТА СТОСТОВО ОТО АЛОСТАТАТ ОТО АЛАГТАТТВ ОТО АЛАГТАТВ ОТО АЛАГТАТВ ОТО АЛАГТАТВ ОТО ОТО ОТО ОТО ОТО ОТО ОТО ОТО ОТО ОТО	TCAIDTOOCT GIAG GAAACHAAA TITO GAAACHAAA TITO GAAACHAAA TITO GAAACHAAA GOTMICLAIT ATO GOTMICLAIT GITO TOCAAACHAA ACTUGITTI AAAC ACCAANTINA TAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA GAACHATTA AAA AAAAAATA AAA GAACHATTA AAAA AAAAAAATA AAAA GAACHATTA AAAA AAAAAATA AAAAAAAAAAAAAAAAAAAA	ТТОВАВ САЛСОВО САЛТОВ ФИЛАНТІ АЛБОСИ АЛБАЛТІ ТЛОТОВ АЛБАЛТІ ТЛОТОВ АЛБАЛТІ ТЛОТОВ АЛБАЛТІ ТОСЛАВ АГАЛТІ ТОСЛАВ АТАЛТОВ ТОСЛАВ АТАЛТОВ ОСТТАС ТРОГЛАВ САЛТОВ САВСТАВ	SCSC AGTT TAAC GITA TTAC AGTT TTAC AGTT TTAC AGTT TTGA TAAAT TCGA AGTA AAT GGTA AAT GGTA GATT TACC AATT AGAA TTTA CATT ATTC COAA TTCA AGTA AASG ATTC	INTERNA INTERNA INTERNA INTERNA INCLASSI INTERNA INTER	

Isoform Gene Proc	luct: cytochrome P450 2E1		
Base numbers are	Relative to start of gene	-	
- RNA splicing: 1 Protein splicing Ribosomal slipp		emove introna. liced to remove inteina. ranslation to generate a progr	
Coding segments	Start base	End base	
Cone 1	1	210	
2		1278	Gap size: 908 bp. Please specify interpreta
323	4223	4372	Gap size: 2944 bp. Please specify interpret
4	4762	4922	Gap size: 359 bp. Please specify interpreta
5	5330	5506	Gap size: 407 bp. Please specify interpreta
6		6635	Gap size: 887 bb. Please specify interprets Gap size: 3165 bb. Please specify interpret
7	9701	9888	Gap size: 500 bp. Please specify interprets
0.	10389	10530	Gap size: 500 Ep. Please specify interpreta
9.	11418	11754	
10.		-	
11.	1	i	-
12	1	1	
12.			
14.	-	1	
15.			
16.	-		
17.			
18.	[		
19.			



### Adding a gene and its encoded protein

Sometimes there is a need to add a gene that is not present in the annotated genome. For the sake of demonstration, we will delete NIES39\_RS03695 and its product, and then recreate them.

- Gene  $\rightarrow$  New gene
- Change classification from "unclassified" to "ORF" (if appropriate)
- Enter gene name (pptX)
- Enter transcription Direction and coordinates (forward; start 779,299; end 779,499
- Enter link to NCBI RefSeq: NIES39\_A08000
- Protein  $\rightarrow$  New
- Macromolecule type  $\rightarrow$  polypeptide
- Gene  $\rightarrow$  pptX
- Name  $\rightarrow$  putative serine/threonine phosphatase
- Enter link to RefSeq: BAI88638.1

# PROTEIN COMPLEXES

#### **Creating Protein Complexes**

• Right-click on a protein and select

Edit  $\rightarrow$  Protein Subunit Structure Editor

 Change "Macromolecule Type" from polypeptide to protein complex

Example: a simple homohexamer

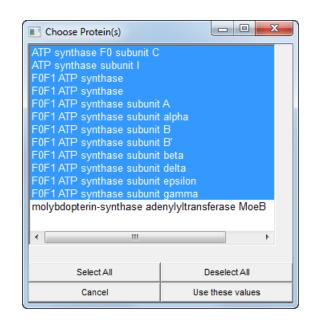
Specify Protein Subunit Structure				×
Protein: serine acetyltransferase				^
Macromolecule Type: protein complex   Number of distinct subunits:	1			
Specific Class(es), if any:				
e.g. A homotetramer counts as 1 gene product, not 4 the number supplied here should match the number of subunits supplied below. For a complex of complexes, check the "Complex?" box below for each subunit				
that is a complex, and enter the number of distinct subunits and the components				
for each. The coefficient can be omitted if it is not known. The Status				
column below tells if a protein already exists or will be created.				
Genes or Subunits:				
Subunit	Complex? Gene or #Subunits	Coefficient	Status eady exists (edit name	
serine acetyltransferase	Gene:   cysE	6 Air	create a new object)	

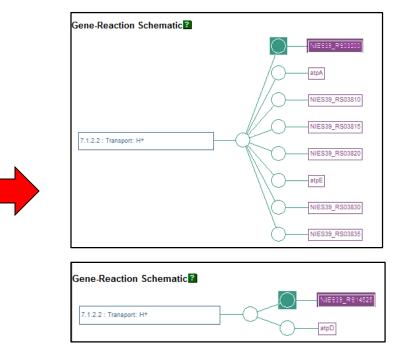
Protein Subunit Structure Editor Add Reaction(s) Add Feature Frame Editor Synonym Editor Relationships Editor Ontology Editor Marvin JS Compound Structure Editor Deprecated Marvin (Java Applet) Compound Structure Editor Import Compound Structure from Molfile Import Compound Structure from ChEBI Export Compound Structure to Molfile Merge Frames Merge Proteins Propagate Enzyme to DB Add Object to SmartTable Add Object to File Export List Duplicate Frame Delete Frame	Protein Editor		^
Add Feature Frame Editor Synonym Editor Relationships Editor Ontology Editor Marvin JS Compound Structure Editor Deprecated Marvin (Java Applet) Compound Structure Editor Import Compound Structure from Molfile Import Compound Structure from ChEBI Export Compound Structure to Molfile Merge Frames Merge Proteins Propagate Enzyme to DB Add Object to SmartTable Add Object to File Export List Duplicate Frame and Edit Create Frame	Protein Subunit Structure Editor		
Frame Editor Synonym Editor Relationships Editor Ontology Editor Marvin JS Compound Structure Editor Deprecated Marvin (Java Applet) Compound Structure Editor Import Compound Structure from Molfile Import Compound Structure from ChEBI Export Compound Structure to Molfile Merge Frames Merge Proteins Propagate Enzyme to DB Add Object to SmartTable Add Object to File Export List Duplicate Frame and Edit Create Frame	Add Reaction(s)		
Synonym Editor Relationships Editor Ontology Editor Marvin JS Compound Structure Editor Deprecated Marvin (Java Applet) Compound Structure Editor Import Compound Structure from Molfile Import Compound Structure from ChEBI Export Compound Structure to Molfile Merge Frames Merge Proteins Propagate Enzyme to DB Add Object to SmartTable Add Object to File Export List Duplicate Frame and Edit Create Frame	Add Feature		
Relationships Editor Ontology Editor Marvin JS Compound Structure Editor Deprecated Marvin (Java Applet) Compound Structure Editor Import Compound Structure from Molfile Import Compound Structure from ChEBI Export Compound Structure to Molfile Merge Frames Merge Proteins Propagate Enzyme to DB Add Object to SmartTable Add Object to File Export List Duplicate Frame and Edit Create Frame	Frame Editor		
Ontology Editor Marvin JS Compound Structure Editor Deprecated Marvin (Java Applet) Compound Structure Editor Import Compound Structure from Molfile Import Compound Structure from ChEBI Export Compound Structure to Molfile Merge Frames Merge Proteins Propagate Enzyme to DB Add Object to SmartTable Add Object to File Export List Duplicate Frame and Edit Create Frame	Synonym Editor		
Marvin JS Compound Structure Editor Deprecated Marvin (Java Applet) Compound Structure Editor Import Compound Structure from Molfile Import Compound Structure from ChEBI Export Compound Structure to Molfile Merge Frames Merge Proteins Propagate Enzyme to DB Add Object to SmartTable Add Object to File Export List Duplicate Frame and Edit Create Frame	Relationships Editor		
Deprecated Marvin (Java Applet) Compound Structure Editor Import Compound Structure from Molfile Import Compound Structure from ChEBI Export Compound Structure to Molfile Merge Frames Merge Proteins Propagate Enzyme to DB Add Object to SmartTable Add Object to File Export List Duplicate Frame and Edit Create Frame	Ontology Editor		
Import Compound Structure from Molfile Import Compound Structure from ChEBI Export Compound Structure to Molfile Merge Frames Merge Proteins Propagate Enzyme to DB Add Object to SmartTable Add Object to File Export List Duplicate Frame and Edit Create Frame	Marvin JS Compound Structure Editor		
Import Compound Structure from ChEBI Export Compound Structure to Molfile Merge Frames Merge Proteins Propagate Enzyme to DB Add Object to SmartTable Add Object to File Export List Duplicate Frame and Edit Create Frame	Deprecated Marvin (Java Applet) Compound Structure Editor	r	
Export Compound Structure to Molfile Merge Frames Merge Proteins Propagate Enzyme to DB Add Object to SmartTable Add Object to File Export List Duplicate Frame and Edit Create Frame	Import Compound Structure from Molfile		
Merge Frames Merge Proteins Propagate Enzyme to DB Add Object to SmartTable Add Object to File Export List Duplicate Frame and Edit Create Frame	Import Compound Structure from ChEBI		
Merge Proteins Propagate Enzyme to DB Add Object to SmartTable Add Object to File Export List Duplicate Frame and Edit Create Frame	Export Compound Structure to Molfile		
Propagate Enzyme to DB Add Object to SmartTable Add Object to File Export List Duplicate Frame and Edit Create Frame	Merge Frames		
Add Object to SmartTable Add Object to File Export List Duplicate Frame and Edit Create Frame	Merge Proteins		
Add Object to File Export List Duplicate Frame and Edit Create Frame	Propagate Enzyme to DB		
Duplicate Frame and Edit Create Frame	Add Object to SmartTable		
Create Frame	Add Object to File Export List		
	Duplicate Frame and Edit		
Delete Frame	Create Frame		
	Delete Frame		
			1

#### Example: defining ATP synthase

Search for ATP synthase subunits: Protein  $\rightarrow$  Search by Substring  $\rightarrow$  ATP synthase

We find 10 subunits arranged into two redundant complexes. Most genes do not have names.





#### Merge the redundant complexes

- Copy the frame ID of the small complex:
   Right-click small complex → Show → Show frame ID
- Merge into the larger complex:
   Right-click large complex → Edit → Merge proteins

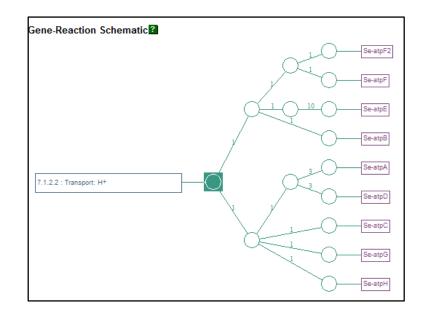
Merge Protein Frames?					
Protein Frame ID to keep CPLX2764-8	Common name: F0F1 ATP synthase				
Protein Frame ID to delete CPLX2764-9	Common name: F0F1 ATP synthase				
Switch Proteins					
OK Cancel					

#### Get some help from MetaCyc

Right click on the reaction and select "Show frame in MetaCyc" Click on the Synechococcus elongatus enzyme

Find out the proper gene name for the different subunits and enter them in the *A. platensis* PGDB

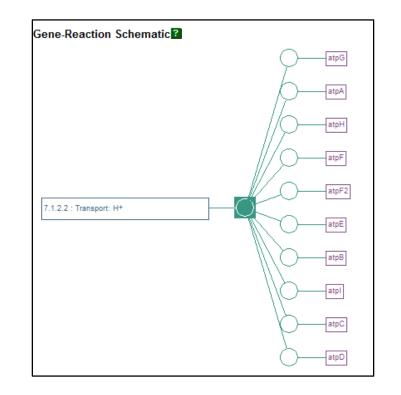
Subunit	gene name
$(F_o sub-complex)$	
а	atpB
b	atpF
b'	atpF2
С	atpG
(F <sub>1</sub> sub-complex)	
α	atpA
β	atpD
γ	atpG
δ	atpH
3	atpC



#### **Correct Atpl annotation**

NIES39\_RS03835-MONOMER was annotated as "ATP synthase subunit I". However, you would notice that Atpl is not part of the MetaCyc complexes. Atpl has been characterized as an accessory protein.

 Change the gene name to *atpl* and the protein name to "ATP synthase accessory factor Atpl"



### Defining the ATP synthase complex (I)

- Right-click on the complex  $\rightarrow$  Edit  $\rightarrow$  Protein Subunit Structure Editor
- Change number of distinct subunits to 2, and clear remaining fields.
- Enter names for the F<sub>o</sub> and F<sub>1</sub> sub-complexes

Specify Protein Subunit Structure		
Name: F0F1 ATP synthase		
Macromolecule Type: protein complex  Number of distinct subunits:	2	
Specific Class(es), if any:		
e.g. A homotetramer counts as 1 gene product, not 4 the number supplied here		
should match the number of subunits supplied below.		
For a complex of complexes, check the "Complex?" box below for each subunit		
that is a complex, and enter the number of distinct subunits and the components		
for each. The coefficient can be omitted if it is not known. The Status		
column below tells if a protein already exists or will be created.		
Genes or Subunits:		
Subunit	Complex? Gene or #Subunits	Coefficient Status
ATP synthase F <sub>o</sub> subcomplex	Gene:	Will be created
ATP synthase F <sub>1</sub> subcomplex	Gene:	Will be created

## Defining the ATP synthase complex (II)

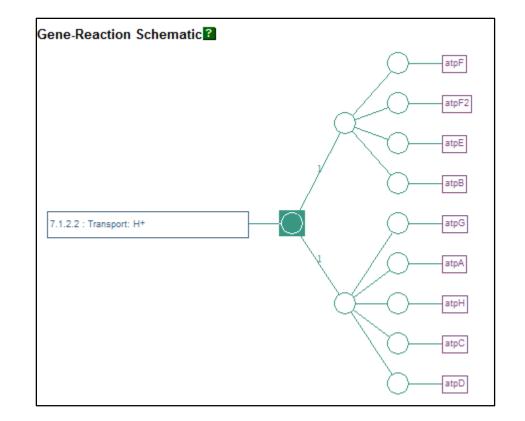
- Mark the two subcomplexes as complexes using the check box.
- Specify 4 subunits for F<sub>o</sub> and 5 subunits for F<sub>1</sub>
- Enter the gene names
- Click OK

(for simplicity we will not worry about coefficients)

Name: F0F1 ATP synthase				
Macromolecule Type: protein complex	ounits: 2			
Specific Class(es), if any:				
e.g. A homotetramer counts as 1 gene product, not 4 the number supple should match the number of subunits supplied below. For a complex of complexes, check the "Complex?" box below for each s that is a complex, and enter the number of distinct subunits and the comp for each. The coefficient can be omitted if it is not known. The Status column below tells if a protein already exists or will be created.	ubunit			
Genes or Subunits: Subunit	Comple	ex? Gene or #Subunit	s Coeffic	cient Status
ATP synthase F <sub>o</sub> subcomplex	<b>v</b>	#Subunits: 4	1	Will be created
F0F1 ATP synthase subunit A		Gene: atpB		Already exists (edit nar to create a new object)
F0F1 ATP synthase subunit B		Gene: atpF		Already exists (edit nar to create a new object)
F0F1 ATP synthase subunit B'		Gene: atpF2		Already exists (edit nan to create a new object)
F0F1 ATP synthase subunit gamma		Gene: atpG		Already exists (edit nar to create a new object)
ATP synthase F <sub>1</sub> subcomplex	•	#Subunits: 5	1	Will be created
F0F1 ATP synthase subunit alpha		Gene: atpA		Already exists (edit nat to create a new object)
F0F1 ATP synthase subunit beta		Gene: atpD		Already exists (edit na to create a new object
F0F1 ATP synthase subunit gamma		Gene: atpG		Already exists (edit nat to create a new object)
F0F1 ATP synthase subunit delta		Gene: atpH		Already exists (edit nat to create a new object)
F0F1 ATP synthase subunit epsilon	F	Gene: atpC		Already exists (edit na
F0F1 ATP synthase subunit epsilon		Gene: atpC		Already exists (edit to create a new obje

#### Complex is defined

Last thing to do is correct the name of the enzyme to  $F_oF_1$  synthase and the activity name to simply ATP synthase



#### **RNA/Protein-RNA complexes**

To create a complex that contains either only RNA molecules or a combination of proteins and RNAs (e.g. a ribosome), simply select the appropriate type of molecule from the "Macromolecule Type" field of the "Protein Subunit Structure" editor

r			The second secon
	Specify Protein S	ubunit Structure	
	Protein: 2-oxoglutarate	e synthase	
	Macromolecule Type:	protein complex 💌	Number of distinct subunits: 4
	Specific Class(e		
1	e.g. A homotetramer o should match the num	protein complex protein-RNA complex per or suburnits supplied per	t 4 the number supplied here ow.
l	For a complex of com	plexes, check the "Complex	?" box below for each subunit
1	that is a complex, and	enter the number of distinct	t subunits and the components
١	for each. The coeffic	ient can be omitted if it is no	t known. The Status
	column below tells if a	a protein already exists or w	rill be created.

### Using citations

- Most editors have citation boxes
- Open the ATP synthase complex in the protein editor
- Add a citation for PMID 33890627

#### **PubMed citations**

- Paste the PubMed ID into a citation box
- Pathway Tools automatically imports the citation when exiting the editor
- Write a summary: "ATP synthase found in the thylakoid membranes of photosynthetic organisms has some unique features not present in other bacterial or mitochondrial systems"
- Use the CITS button to move the citation to the summary

#### Using non-Pubmed citations

You can type Pubmed IDs directly into the summary using CITS. Other types of citations must be first entered into a citation box as described earlier. Once they have been created, they can be used in the summaries

- Enter an ID in the form Krah10 in a citation box, invoke editor by clicking out of the box. Click on "Search or Create Publication Frame".
- If you have a DOI number, enter it and click outside the DOI ID box, and it will be retrieved automatically (e.g. 10.1016/j.jmb.2009.10.059)
- If there is no DOI, type in the details.
- Click OK to close the editor

To make corrections invoke the Publication Editor by right-clicking on a citation and selecting Edit  $\rightarrow$  Publication Editor

d ID:	AGRICOLA ID:	DOI ID: 10.1016/j.jmb.2009.10.0	59
On the Struc	ture of the Proton-Binding Site	in the Fo Rotor of Chloroplast ATP Synth	*
rs (surname first	): 1. Krah A	2. Pogoryelov D	
	3. Meier T	4. Faraldo-Gomez JD	_
	5.	6.	_
	,	,	_
ce: Journal of	Molecular Biology 395(1);20-27		Year: 2010



#### Writing Summaries

By writing summaries for enzymes and pathways you can turn your PGDB into a resource that integrates and summarizes the current knowledge about your organism.

Summaries should incorporate references to the literature, using the CITS button.

Using internal hyperlinks to other database objects is extremely useful. See for example

https://biocyc.org/SYNWH8102/NEW-IMAGE?type=ENZYME&object=CPLX1YI0-83

#### **Using Internal Hyperlinks**

- Internal hyperlinks are entered by using the |FRAME| button
- Clicking the button opens a list of the items in the History.
   Select the right one, and the hyperlink is created
- If the History list becomes too long, you can reset it by selecting Tools → History → Clear
- Alternatively, you can skip the history list by clicking the FRAME button a second time, which will insert an empty link. Print the frame ID of an object to the lisp window (right-click → Show → Show frame ID), copy it to the clipboard, and past it into the link.
- To modify what is shown in the summary, use this format |FRAME: frame-ID "type text here"|.
- You can use hyperlinks to MetaCyc objects.

#### Example for a summary with hyperlinks

EC 2.7.2.4, aspartate kinase, catalyzes the committed step in the pathways that ultimately lead to the synthesis of the amino acids lysine, threonine and isoleucine.

|FRAME: EC-2.7.2.4|, catalyzes the committed step in the pathways that ultimately lead to the synthesis of the amino acids |FRAME: LYS|, |FRAME: THR| and |FRAME: ILE|.



EC 2.7.2.4, aspartate kinase, catalyzes the committed step in the pathways that ultimately lead to the synthesis of the amino acids L-lysine, L-threonine and L-isoleucine.

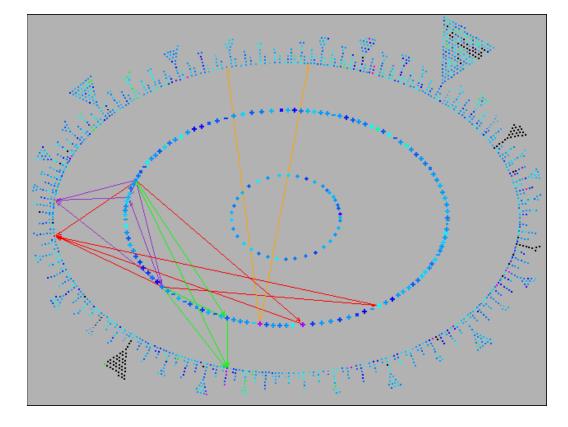
#### Copying MetaCyc comments

Pathways show up with the MetaCyc summary. If you want to modify a summary so it is applicable to your organism, you can copy the MetaCyc summary and alter it.

To copy a summary from MetaCyc:

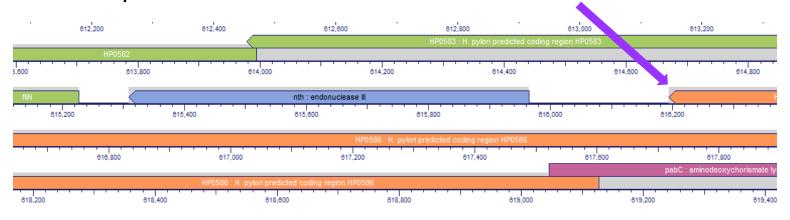
- 1. Type (dev) in the listener pane to activate MetaCyc editors
- 2. Right-click the pathway's name in your PGDB, select Show  $\rightarrow$  Show frame in MetaCyc
- Open the MetaCyc pathway in the Pathway Info Editor and copy the text
- 4. Go back to the pathway in your PGDB, open the editor and paste

#### **Curating Regulation**



### **Transcription Units**

Transcription units are predicted by the PathoLogic tool "predict transcription units".



Right-click the gray area and select Edit  $\rightarrow$  Transcription Unit Editor. You can:

- Modify the included genes
- Add promoter information
- Specify sigma factor
- Add several different experimental evidence codes

#### Example for regulatory data

Let's make some (unbased) assumptions:

- 1. The *atpE*, *atpB*, and *atpI* genes are part of the transcription unit that contains the 5 genes upstream.
- 2. This TU is controlled by the sigma70 factor (NIES39\_RS27630-MONOMER)
- 3. There is a promoter (atplp) 85 bases before the atpl gene
- 4. Transcription factor (TF) NIES39\_RS27155-MONOMER activates transcription, but only when bound to ADP

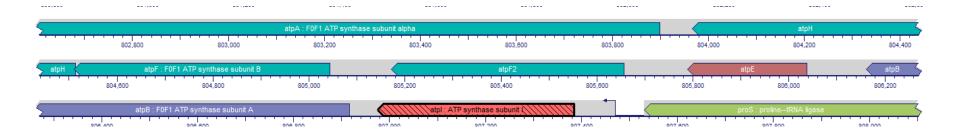
#### Modifying the transcription units

First we need to remove the incorrect TUs. Right-click on the Tus of *atpE*, *atpB*, and *atpI*, respectively, and select Edit  $\rightarrow$  Delete frame (can be done from the gene page).

Next, right-click on the TU that needs expansion and select Edit  $\rightarrow$  Transcription Unit Editor. Add the names of the three genes to the list of genes.

In the Promoter box type "atplp", and enter -85 in the box "distance from the first gene".

From the list of sigma factors, select the sigma70 family factor.



#### Adding transcription factor (TF) interactions

- Click on TU and select Edit  $\rightarrow$  Create Regulatory Interaction
- In the window that pops up, select the atplp promoter
- Paste the TF frame ID (NIES39\_RS27155-MONOMER) in the "Protein" box
- Select "Activator" from the "Function" box
- Enter ADP for small molecule, and select "Active" in the box below
- Add evidence code and references
- Optional: define relative distance from transcription start site
- The information can be edited in the future by right-clicking on the green box and selecting Edit  $\rightarrow$  Regulatory Interaction Editor



#### **Specifying Terminators**

# Right-Click on the TU Edit $\rightarrow$ Edit Terminators

Enter Terminator Data			×
Left End Position Right End Position	Sequence	Rho-dependent?	Citation
OK Cancel			

#### **More Regulatory Interactions**

Pathway Tools supports additional types of regulatory interactions including:

- Attenuation
- Regulation of translation
  - RNA-mediated
  - Protein-mediated
  - Small molecule-mediated
- Regulated protein or mRNA degradation

When starting the Regulatory Interaction Editor, click the "Type of regulation" button at the top to select these types of interactions.

#### PathoLogic PGDB refinement tools

PathoLogic contains a number of tools that require curator interaction, that can add content and improve the PGDB quality

atabase Build 🛛	Refine
rganism:	Assign Probable Enzymes
D: TEST	Assign Modified Proteins
lame: A. plate	Protein Complex Inference, Automated
Strain: NIES-39	Create Protein Complexes, Manually
	Re-Run Name Matcher
atus: Built	Rescore Pathways
	Predict Transcription Units
netic Elemer C 016640	Import Curated MetaCyc Data
C_010040	Update Cellular Architecture
te: A log of all mess	Transport Inference Parser
ioLogic console win Jsers\Ron\AppData vay Tools\PathoLogi sole-Log_2022-03-0	Update Overview
	Pathway Hole Filler
	Run Consistency Checker

#### Assign modified proteins

In the previous webinar we discussed the fact that proteins that participate in MetaCyc reactions are all classes.

This tool enables linking the protein instances in the PGDB with these protein classes

a [ThiS sulfur-carrier protein]						
	NIES39_RS02995 : MoaD/ThiS family protein NIES39_RS07415 : MoaD/ThiS family protein thiS : thiamine biosynthesis protein ThiS					
Show Candidate(s)	hisH : imidazole glycerol phosphate synthase subunit HisH hisF : imidazole glycerol phosphate synthase subunit HisF	Add Candidate by Gene ID:				
	<►	Þ.				
Show Reaction(s)	Show Reaction(s) RXN-9789 : a [ThiS sulfur-carrier protein] + ATP + H+ -> a carboxy-adenylated-[ThiS sulfur-carrier protein] + diphosphate					

#### Create protein complexes (manual tool)

Pathway Tools contains an automated protein complex creation tool. However, that tool only accepts complexes whose subunits are encoded by neighboring genes.

This manual tool allows a curator to inspect possible additional complexes and accept or reject them.

#### Group Subunits into Complexes for Reaction RXN0-2381 (3 out of 133) Group Subunits into Complexes for Reaction RXN0-2381 Note: None of these genes are neighbors on the replicon To add an additional protein subunit that is not included in one of the lists below, navigate to it in the Pathway Tools Navigator window, to add it to the history list, and then press the "Add protein subunit to list ..." button below the list. Complex 1: NIES39 RS13305 (trpA) : tryptophan synthase subunit alpha NIES39 RS27525 (trpB) : tryptophan synthase subunit beta 111 Add protein subunit to list for Complex 1 Add Complex 111 Make Complex(es) Skip Stop

### Transport Inference Parser (TIP)

TIP attempts to identify transporters in the genome based on their annotation and creates the most likely transport reactions that describes their activity.

It also creates complexes when it identifies a set of genes likely to encode a complex.

Transport Inference Parser Dialog			
Protein ID: CPLX2764-18			
Function Name: putative Fe3+ ABC transporter			
Components: Subunit NIES39_RS04740-MONOMER Function Name: iron ABC transporter permease Subunit NIES39_RS04745-MONOMER Function Name: iron ABC transporter substrate-binding protein			
Modify Components			
Energy coupling: ATP			
Substrate: Fe3+			
Reaction: Fe3+[periplasm] + ATP + H2O -> Fe3+[cytosol] + ADP + phosphate + H+			
Retract reaction Edit reaction			
Accept Reject Leave status unchanged			
· · · · · · · · · · · · · · · · · · ·			

#### Pathway Hole filler (PHF)

The Pathway Hole filler attempts to identify missing enzymes in pathways based on sequence similarity to enzymes that have been assigned to these reactions in MetaCyc. This can overcome shortcomings in the annotation pipeline.

The PHF runs in three steps.

Step 1: Identify pathway holes, gather sequence data and build a training data set (takes 10-20 minutes) -automatic
Step 2: Identify candidate genes – automatic
Step 3: This is the step where the curator needs to go over the suggestions and decide which ones to accept - manual



The tool correctly identifies a missing enzyme for the phylloquinol biosynthesis pathway, which has been misannotated.

Choose Holes to Fill in KB		×		
Instructions: If you click on the name or description of any biological ob To consider only high-probability hole-filling candidates, ple Minimum probability cutoff (Range: 0.0000000 to 1.000000				
		Fill hole with top candidate?		
Holes/Reactions	Top candidate	Set all to Yes Set all to No		
Pathway phylloquinol biosynthesis:				
EC# EC-2.1.1.329: S-adenosyl-L-methionine + demethylphylloquinol -> S-adenosyl-L-homocysteine + phylloquinol + H*	ubiE/NIES39_RS19045 P = 0.9448 Show all 10 candidates	<ul> <li>No</li> <li>Yes, by adding function</li> <li>Yes, by replacing function</li> </ul>		
•	III	•		
OK Cancel Save Selections to DB				

## And now, brought to you by Chitty Chitty Bang Bang and SRI International...

# If your PGDB started out all smooth and shiny...





#### ...but now it looks like this

#### ...then it's time for an overhaul!

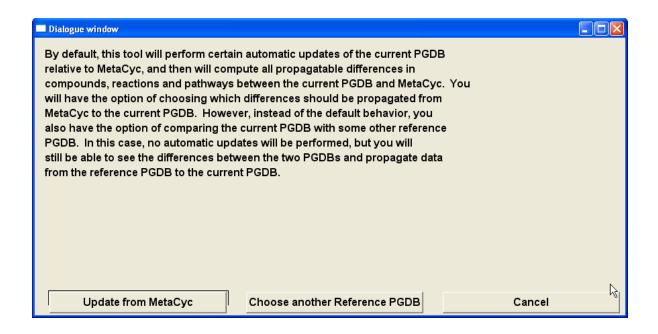
Tools for updating an aging PGDB

- Propagate updates from Reference DB (MetaCyc)
- Re-run the name matcher
- Rescore pathways
- Run the consistency checker



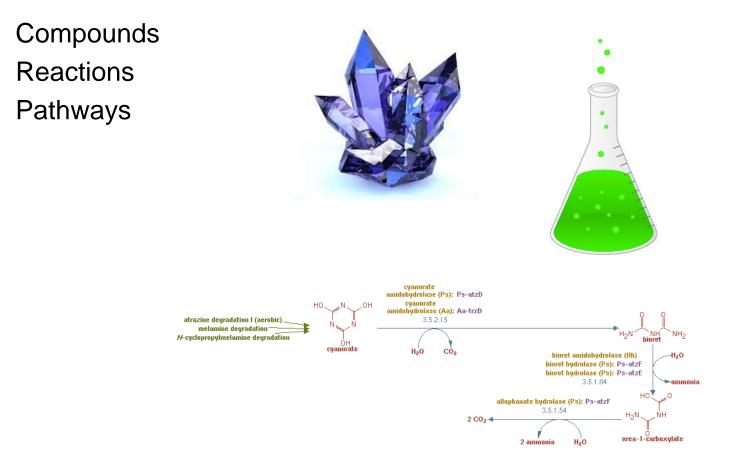
#### Propagating updates from a reference PGDB

Invoke from the Tools menu (Propagate MetaCyc Data Updates) If your PGDB was created using a different reference PGDB, you can select it instead of MetaCyc



#### Propagating data updates

Data updates are broken into three sections:



## Propagating compound data

# For compounds, the software looks for differences in chemical structures and in the data stored in the different slots

Compounds			
33 Compounds have structures in MetaCyc but not in HpyCyc.	Select for Update	Propagate All	
537 Compounds have structure differences between MetaCyc and HpyCyc.	Select for Update	Propagate All	
5 Compounds have differences in slot N+1-NAME between MetaCyc and HpyCyc.	Select for Update	Propagate All	
5 Compounds have differences in slot N-1-NAME between MetaCyc and HpyCyc.	Select for Update	Propagate All	
4 Compounds have differences in slot N-NAME between MetaCyc and HpyCyc.	Select for Update	Propagate All	
2 Compounds have differences in slot OVERVIEW-NODE-SHAPE between MetaCyc and HpyCyc.	Select for Update	Propagate All	
42 Compounds have differences in slot COMMON-NAME between MetaCyc and HpyCyc.	Select for Update	Propagate All	
7 Compounds have differences in slot CITATIONS between MetaCyc and HpyCyc.	Select for Update	Propagate All	Merge All
2 Compounds have differences in slot GIBBS-0 between MetaCyc and HpyCyc.	Select for Update	Propagate All	
540 Compounds have differences in slot DBLINKS between MetaCyc and HpyCyc.	Select for Update	Propagate All	Merge All
141 Compounds have differences in slot SYNONYMS between MetaCyc and HpyCyc.	Select for Update	Propagate All	Merge All
9 Compounds are present in HpyCyc but not in MetaCyc.	Examine		

#### Inspecting differences

When you click the "select for update" button, you can review the differences and decide what to do for each case.

Differences in slot CITATIONS				
Select All for Propagation	Select All for Merge Unselect	t All	Update Selected	
Propagate? Merge?	Object H	pyCyc Valu	e MetaCyc Value	
4-AMINO-BUT 4-aminobutyr		NIL	Steward49	Show Object Displays
	-phosphoribitylamino)uracil	NIL	("11889103" "18245297")	Show Object Displays
CPD-9923 (1 <i>R</i> ,6 <i>R</i> )-6-hy -carboxylate	droxy-2-succinylcyclohexa-2,4-diene-1	NIL	18284213	Show Object Displays
CPD-9924 2-succinyl-5- -carboxylate	enolpyruvyl-6-hydroxy-3-cyclohexel/g-1	NIL	18284213	Show Object Displays
CPD-9925	/-2-naphthoyl-CoA	NIL	11153266	Show Object Displays
CPD-7670 dimethylsulfid	e	NIL	CHARLSON87	Show Object Displays
		NIL	10382261	Show Object Displays

#### Propagating reaction data

For reactions, the software looks for differences in the reaction equation, as well as in the data stored in the different slots

Reactions		
307 Reactions have equation differences between MetaCyc and NpyCyc.	Select for Update	Propagate All
46 Reactions have differences in slot COMMON-NAME between MetaCyc and HpyCyc.	Select for Update	Propagate All
16 Reactions have differences in slot EC-NUMBER between MetaCyc and HpyCyc.	Select for Update	Propagate All
6 Reactions have differences in slot PREDECESSORS between MetaCyc and HpyCyc.	Select for Update	Propagate All
79 Reactions have differences in slot OFFICIAL-EC? between MetaCyc and HpyCyc.	Select for Update	Propagate All
1 Reactions have differences in slot SPONTANEOUS? between MetaCyc and HpyCyc.	Select for Update	Propagate All
11 Reactions are present in HpyCyc but not in MetaCyc.	Examine	

#### Objects not present in the reference database

When the software finds objects in the PGDB that are missing from the reference database, you can click the "Examine" button next to it to see the details.

The software would try to find merge candidates for these objects

F	RXN-3	3781	Show		
r	malat	e = oxaloacetate			
	F	Merge with RXNI-3 (malate + menaquinone-8 -> oxaloacetate + menaquinol) from HpyCyc	Show		
	Γ	Merge with MALATE-DEH-RXN (malate + NAD <sup>+</sup> = oxaloacetate + NADH) from HpyCyc	Show		
	Γ	Merge with MALATE-DEHYDROGENASE-NADP+-RXN (malate + NADP <sup> = oxaloacetate + NADPH + H<sup>+</sup>) from MetaCyc</sup>	Show		
	Γ	Merge with MALATE-DEHYDROGENASE-ACCEPTOR-RXN (malate + an oxidized electron acceptor = oxaloacetate + a reduced electron acceptor) from HpyCyc	Show		
	Γ	Merge with MALOX-RXN (malate + O <sub>2</sub> = oxaloacetate + hydrogen peroxide) from MetaCyc	Show		
	Γ	Merge with LACTATE-MALATE-TRANSHYDROGENASE-RXN (oxaloacetate + L-lactate = pyruvate + malate) from MetaCyc	Show		
	PABS	YNMULTI-RXN			
	L-glutamine + chorismate = p-aminobenzoate + L-glutamate + pyruvate				
	No	nerge candidates were found for this object			
Sel		for Deletion Unselect All Delete/Merge Selected			

## Propagating pathway data

For pathways, the software looks for differences in the topology of the pathway, as well as in the data stored in the different slots

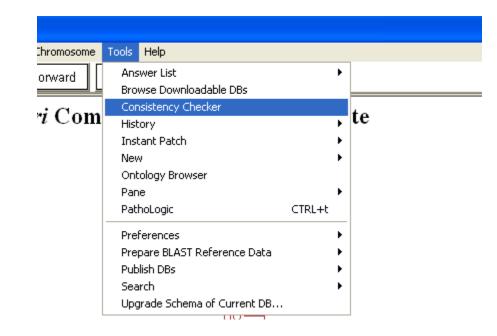
When pathways are present in your PGDB but not in the reference PGDB, it may be for two reasons: either you created them (in which case you would probably want to keep them), or they were deemed incorrect or redundant in MetaCyc, in which case you would want to delete them.

To make life easier: when modifying pathways in your PGDB, change the frame ID!

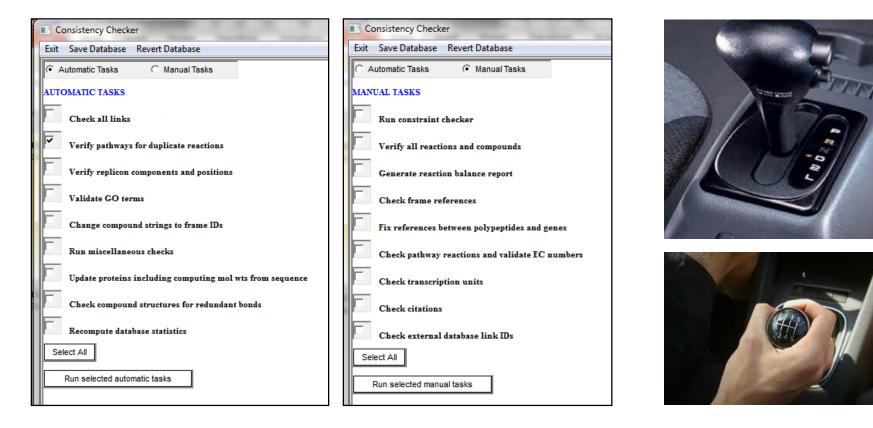
Pathways			
36 Pathways have topological differences between MetaCyc and HpyCyc.	Select for Update	Propagate All	
2 Pathways have differences in slot HYPOTHETICAL-REACTIONS between MetaCyc and HpyCyc.	Select for Update	Propagate All	
10 Pathways have differences in slot SYNONYMS between MetaCyc and HpyCyc.	Select for Update	Propagate All	Merge All
118 Pathways have differences in slot CREDITS between MetaCyc and HpyCyc.	Select for Update	Propagate All	Merge All
20 Pathways have differences in slot COMMON-NAME between MetaCyc and HpyCyc.	Select for Update	Propagate All	
11 Pathways are present in HpyCyc but not in MetaCyc.	Examine		

#### The consistency checker

Consistency checking should be performed routinely (every few months), and problems should be addressed



#### Automatic and manual tasks



- I recommend running the automatic tasks first
- I recommend running individual tasks one at a time.
- When you mouse over a task's name, you will see documentation for that particular task in the bottom window pane.

#### Consistency checker output

 The output appears on the right pane but is also saved into a text file in the reports directory. The name and location of the file are printed at the end of the output.



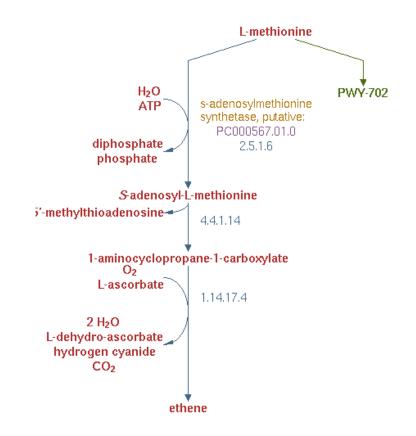
==Done checking all the links==

The report from this consistency checker run can be found at

C:\Program Files\Pathway Tools\ptools-local\pgdbs\registry\hpycyc\13.1\reports\consistency-checker-report-2009-08-13\_11-24-56.txt

#### Automatic tasks: check all links

This tool looks at: Inverse links (compoundreaction, gene-protein, etc.) Pathway links Ghost reactions in pathways Pathways included in other pathways



===== Checking and removing any values from PATHWAY-LINKS that point to nonexistent frames ==== Removing link from pwy **PWY-5901** to nonexistent pwys (ENTBACSYN-PWY)

#### Automatic tasks: check all links

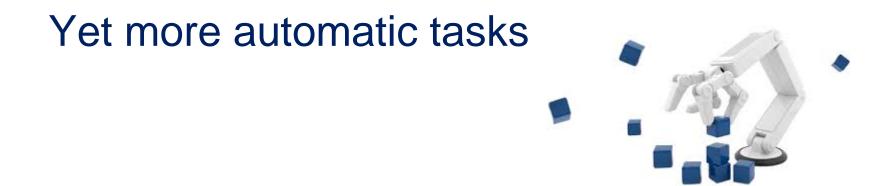
Warnings are not necessarily errors but should be checked.

For example, PWY-21 is completely redundant to P142-PWY and should probably be deleted.

F.
Warning:MET-SAM-PWY is completely contained within PWYI-3 but is not listed in the SUB-PATHWAYS slot
Warning: P142-PWY is completely contained within PWY-21 but is not listed in the SUB-PATHWAYS slot
Warning: PWY-5600 is completely contained within PWY-21 but is not listed in the SUB-PATHWAYS slot
Warning:GLYCOLYSIS is completely contained within ANAEROFRUCAT-PWY but is not listed in the SUB-PATHWAYS slot
Warning: PWY-5485 is completely contained within FERMENTATION-PWY but is not listed in the SUB-PATHWAYS slot
Warning: PWY-21 is completely contained within P142-PWY but is not listed in the SUB-PATHWAYS slot
Warning: PWY-21 is completely contained within PWY-5600 but is not listed in the SUB-PATHWAYS slot
Warning: PWY-5484 is completely contained within GLYCOLYSIS but is not listed in the SUB-PATHWAYS slot

#### More automatic tasks

- Verify pathways for duplicate reactions
- Verify replicon components and positions: ensures all genes exist, sorts based on position.
- Validate GO terms: updates the GO terms using the latest version of GO-KB, removes obsolete ones.
- Change compound names to string IDs: mostly applies to legacy data, where enzyme regulators may have been entered as text strings.



- Run miscellaneous checks: formatting glitches in names, validity of superpathways, clears values of computed slots, deletes temporary frames created by breaks when the pathway editor runs
- Update proteins: molecular weights recalculated from sequence
- Check compound structures for redundant bonds

#### Automatic tasks: recompute database statistics

#### Updates the numbers on the home page

	He	licobacter	pylori		
	Stra	in: 26695 HpyCyc v	rersion: 13.1		
Ge	nerate Pathway Evid	ence Report	Generate Pathway	Hole Report	
Authors: Suzanne Pale SRI Interns	ational; Peter D. Kar	p, SRI International			
- -					
Citations: [Tomb, 1997; Marais, 19	99]				
Replicon	Total Genes	Protein Genes	RNA Genes	Pseudogenes	Size (bp)
26695 Chromosome	1609	1566	43	0	1,667,867
		Pathways:	143		
		Enzymatic Reactions:	671		
		Transport Reactions:	29		
		Polypeptides:	1598		
		Protein Complexes:	29		
		Enzymes:	330		
		Transporters:	33		
		Compounds:	553		
		Transcription Units:	817		
		tRNAs:	38		

#### Manual tasks: the constraint checker

This tool usually requires the most time and effort for correcting the problems.



Flags constraints issues. For example, if a slot is supposed to contain only compound frame IDs, but a different type of frame is listed among its values, the constraint checker identifies and flags the offensive value.

The opposite is true as well: the checker will flag that compound as present in a slot of a frame that is not supposed to have such a value.

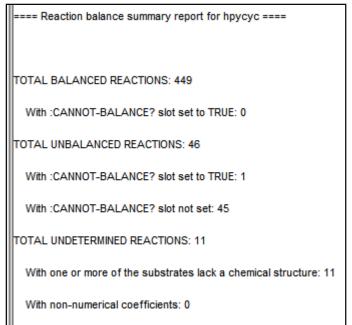
(this means errors are often listed multiple times, under different frames)

The checker also flags cardinality violations. For example, cases where more than one values are present in a slot that is only allowed to have a single value.

#### More manual tasks

- Verify all reactions and compounds: finds orphan enzymatic reaction frames (missing a protein, a reaction, or both); finds orphan reactions that are not associated with any other objects, looks for duplicate compounds.
- Generate reaction balance report





#### Frame references errors

Frame AGMATHINE. is referenced in a  FRAME:   construct, but
des not exist either here or in MetaCyc or in EcoCyc. It is referenced in the
following places:
Frame: PWY0-1299
Slot: COMMENT



Looking at that pathway's comment, we find that the FRAME construct is missing the last bar.

ginine-dependent acid resistance system which couples gmatine antiporter, AdiC, with arginine decarboxylase, AdiA. nal [FRAME: ARG] for internal [FRAME: AGMATHINE. Arginine II arginine is decarboxylated by AdiA to agmatine, releasing rith a proton. Agmatine is then exported through AdiC.

#### More manual tasks

- Fix references between polypeptide and genes: adds the gene value to modified proteins that miss it, adds a capitalized gene name to the synonyms list, scans that list for duplicates, flags orphan genes and proteins.
- Check pathway reactions and validate EC numbers: checks the PREDECESSORS slot of pathway frames, flags references to deleted and transferred EC numbers.
- Check transcription units: looks for invalid frames, transcription units with no genes, with genes in different directions, etc.

#### Even more manual tasks

- Check citations: tries to find formatting problems, downloads
   PubMed citations that have not been imported, provides
   statistics.
- Check external database link IDs: flags frames that are linked to the same external DB entry by links that are supposed to be unique.
- Check HTML tags: looks for formatting errors in HTML within comments.

# And when you finish, take pride at your newly renovated PGDB!



#### Homework

Perform the following exercises at your pleasure

Exercise 2: Assigning enzymes and creating protein complexes

- Assigning enzymatic activities to proteins
- Defining protein complexes
- Creating a publication frame
- Exporting a pathway to a file