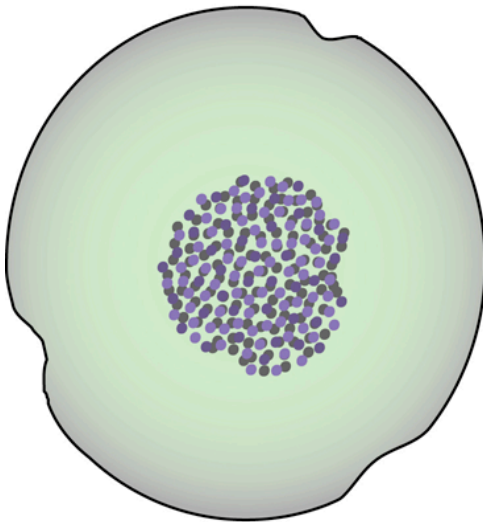


# **Possible Bacterial Origin of the NPC**

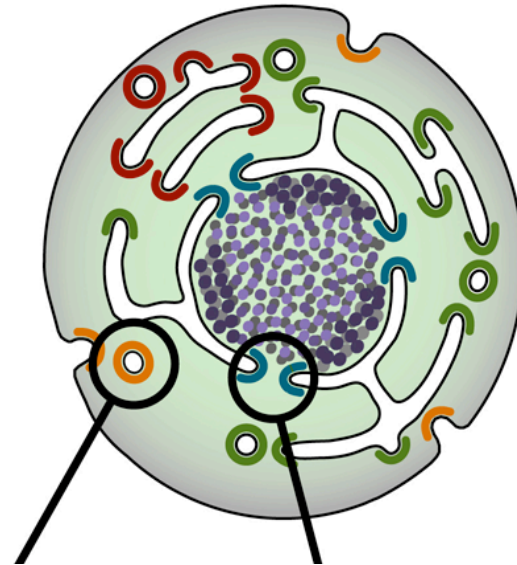
Damien Devos  
Salilab GM 07 Nov 2005

# Eukaryotic evolution

Prokaryote



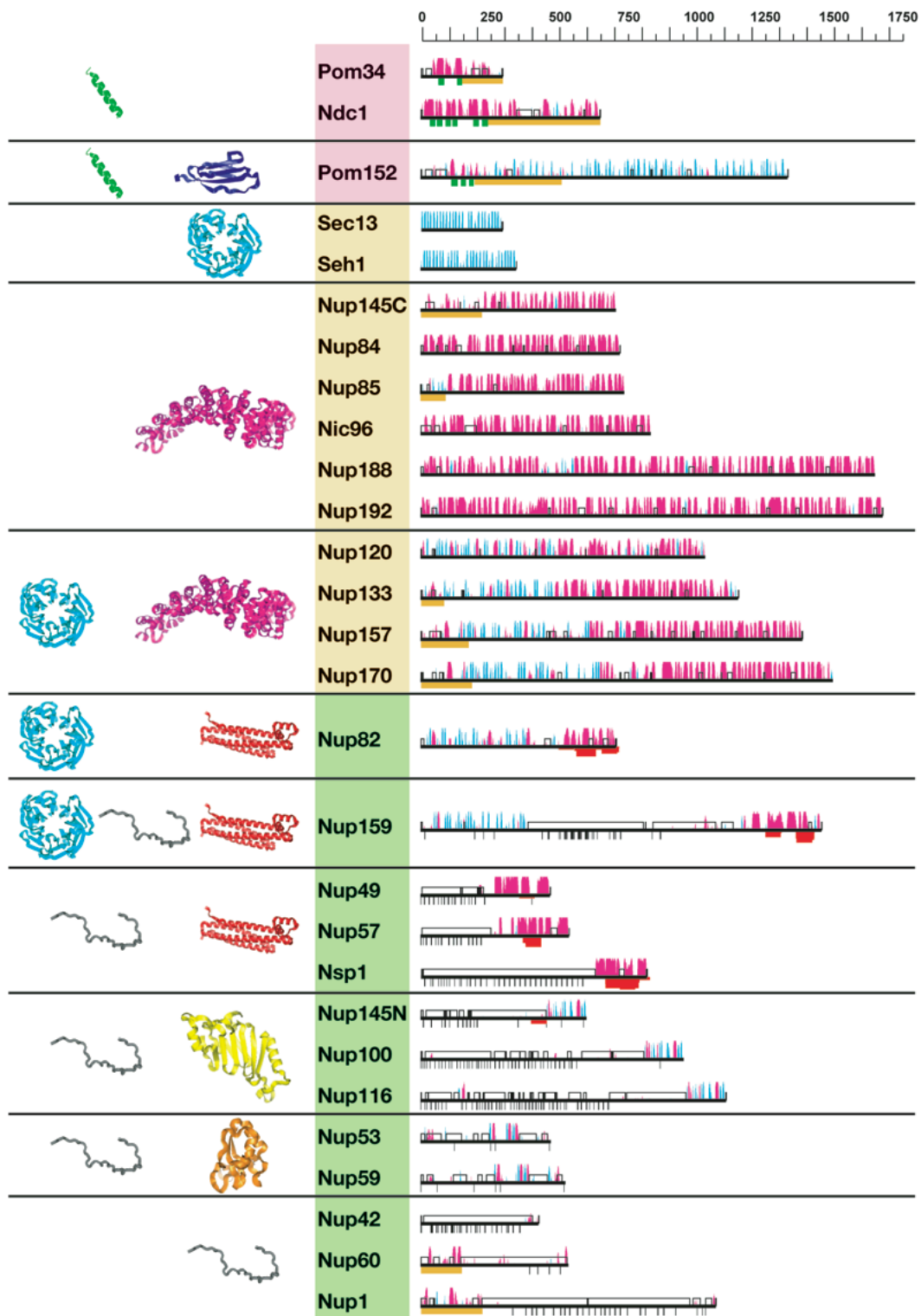
Modern Eukaryote



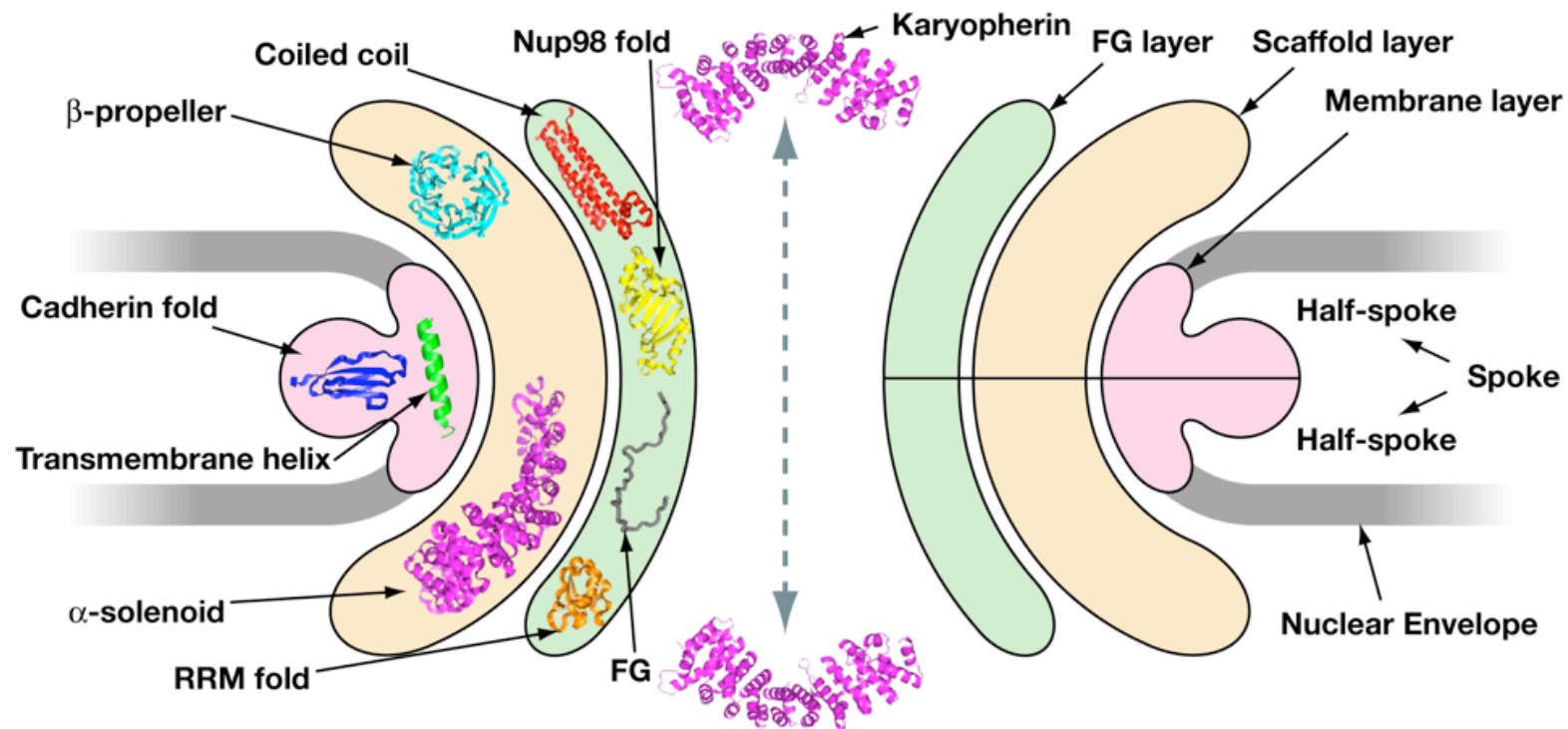
?

Complex internal compartmentalization

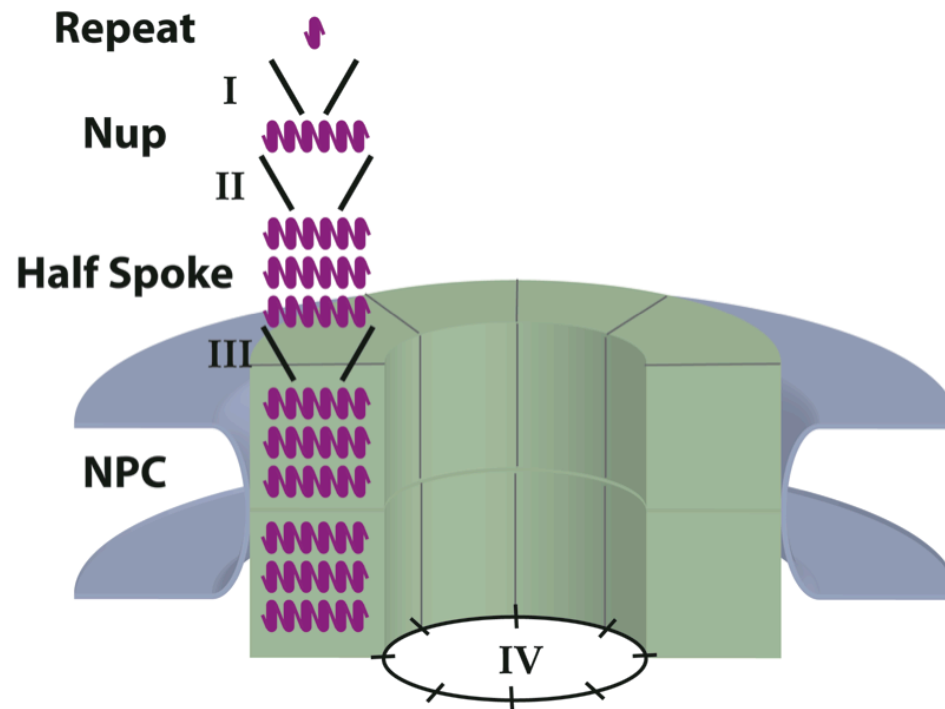
How could such a complicated system evolve in organisms with no analogous transport system?



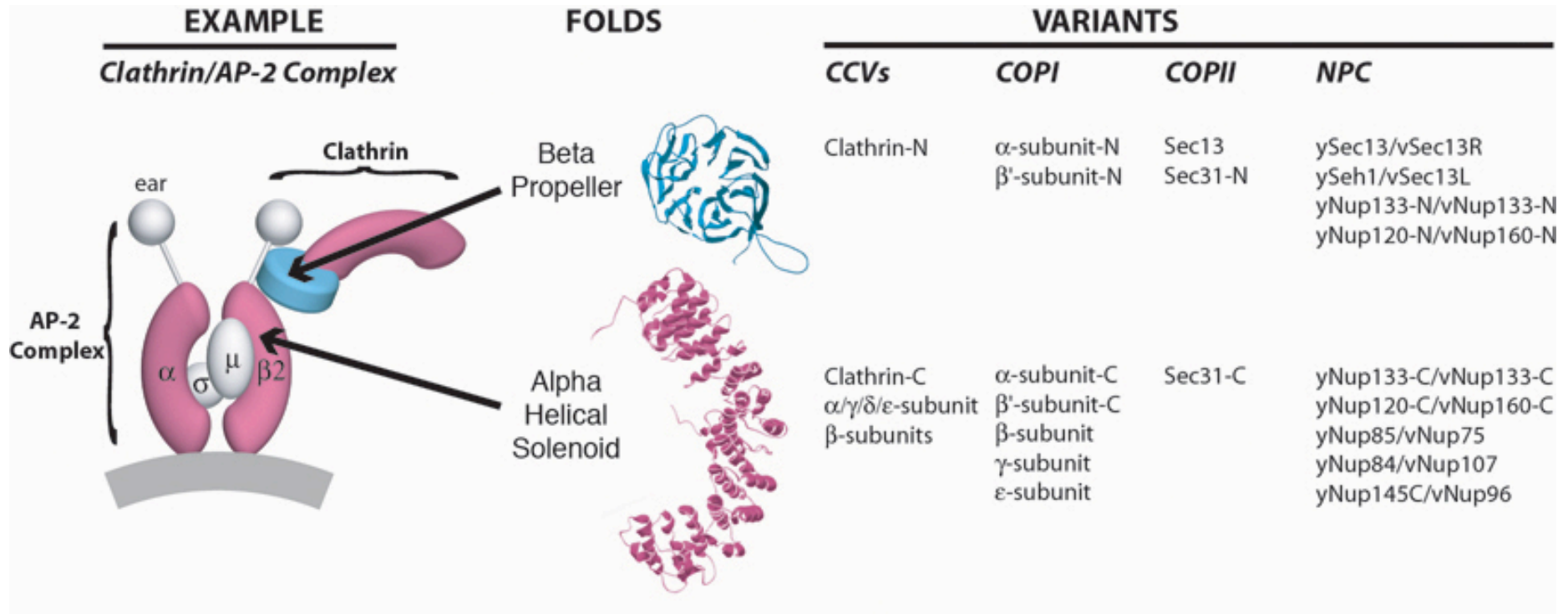
# NPC is organized in 3 layers



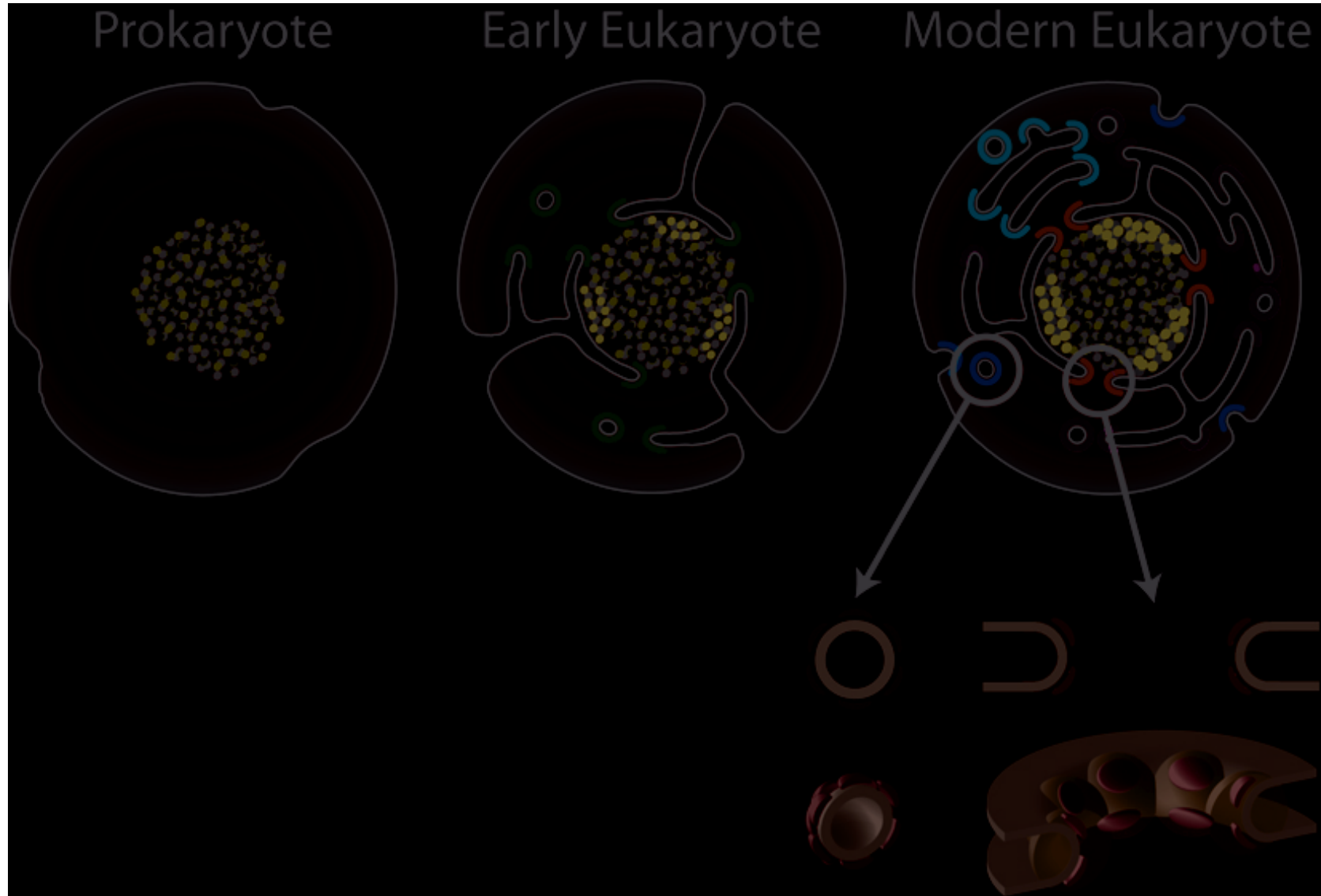
# Hierarchical organization of the NPC



# NPC and Coated Vesicle



# Eukaryote evolution



# The Proto-coatomer hypothesis

- Proteins with the proto-coatomer architecture are only common protein between components of the eukaryotic endomembrane system
- => A search of proto-coatomer architecture in other system might reveal clues to its origin!



# Homology-based proto-coatomer search

- ModBase & SuperFamily
- => Only eukaryotic proteins detected
- No clues about origin

Plos Biology 2004

# HMM search

- HHsearch (Soding, 2004 Bioinformatics)
- All genomes (euk, bact, archea)
- $\beta$ -propeller &  $\alpha$ -solenoid detection

```
#!/bin/csh
#
#$ -S /bin/csh
#$ -o codine-hhsbactleft-logs
#$ -e codine-hhsbactleft-logs
#$ -cwd
#$ -r y
#$ -j y
#$ -l dival=1G,diva3=1G,scratch=1G,i386=true
#$ -p -4
#$ -N hhs_bactleft
#$ -t 1-55676
set tasks = ( CJEJ-221-01-001400.fSFLE-457-01-004067.f SFLE-457-01-004068.f ....)
set input = $tasks[$SGE_TASK_ID]

date
hostname

setenv MODPIPEBASE /dival/home/ModPipe/SVN

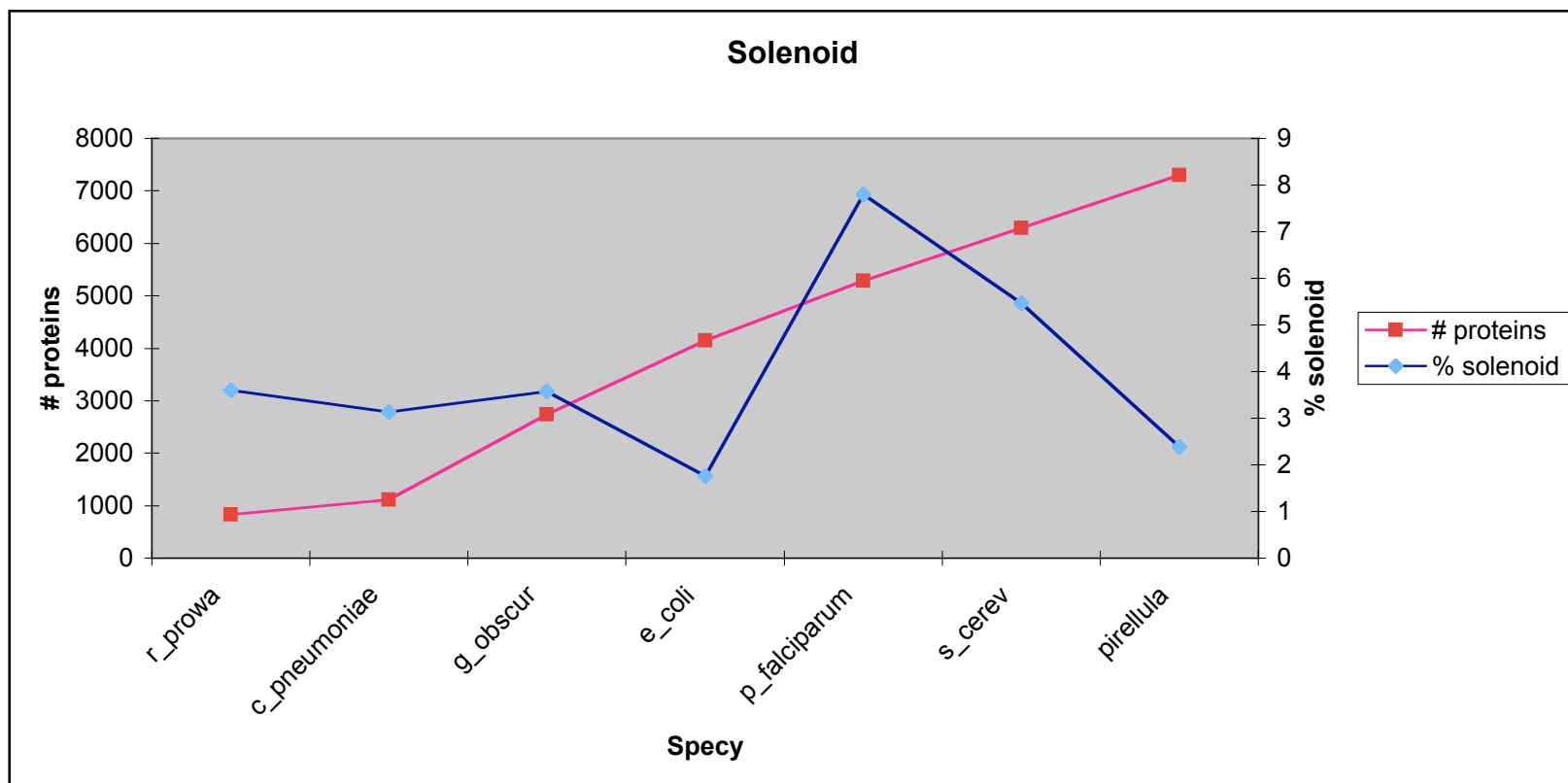
set org = `echo ${input} | awk -F"." '{print $1}`
echo "${input} is input"
echo "${org} is org"

set tmpdir = /scratch/damien/hmm-${input}
set panfs = /panfs/salilab/home/damien
mkdir -p ${tmpdir}
cd ${tmpdir}/
mkdir -p ${panfs}/hhsearch/bacteria/${org}

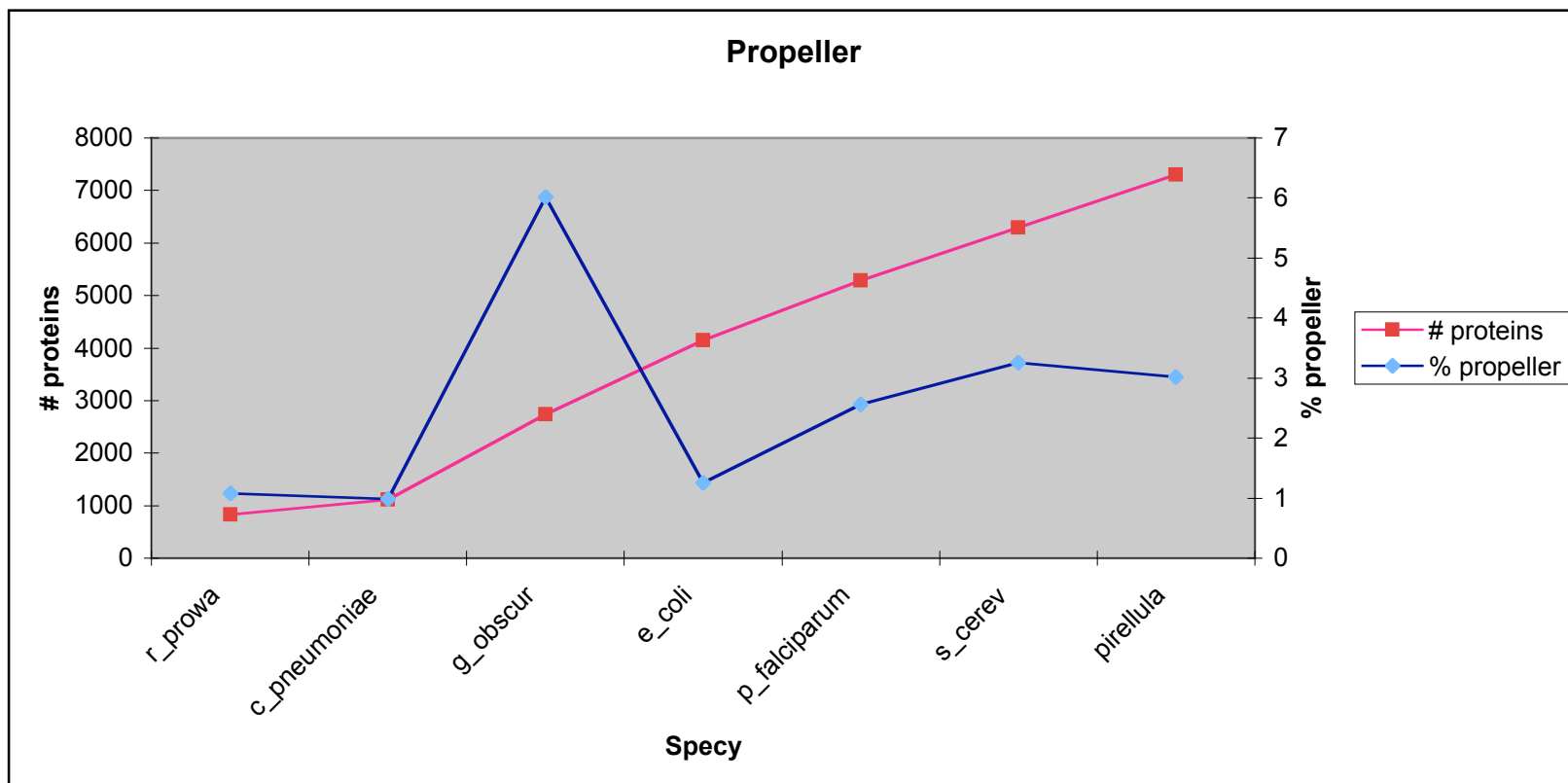
$MODPIPEBASE/aux/CopyFile -i ${panfs}/hhsearch32 -o ./hhsearch32
$MODPIPEBASE/aux/CopyFile -i ${panfs}/genomdb/bacteria/${org}/${input} -o ./${input}

chmod +x hhsearch32

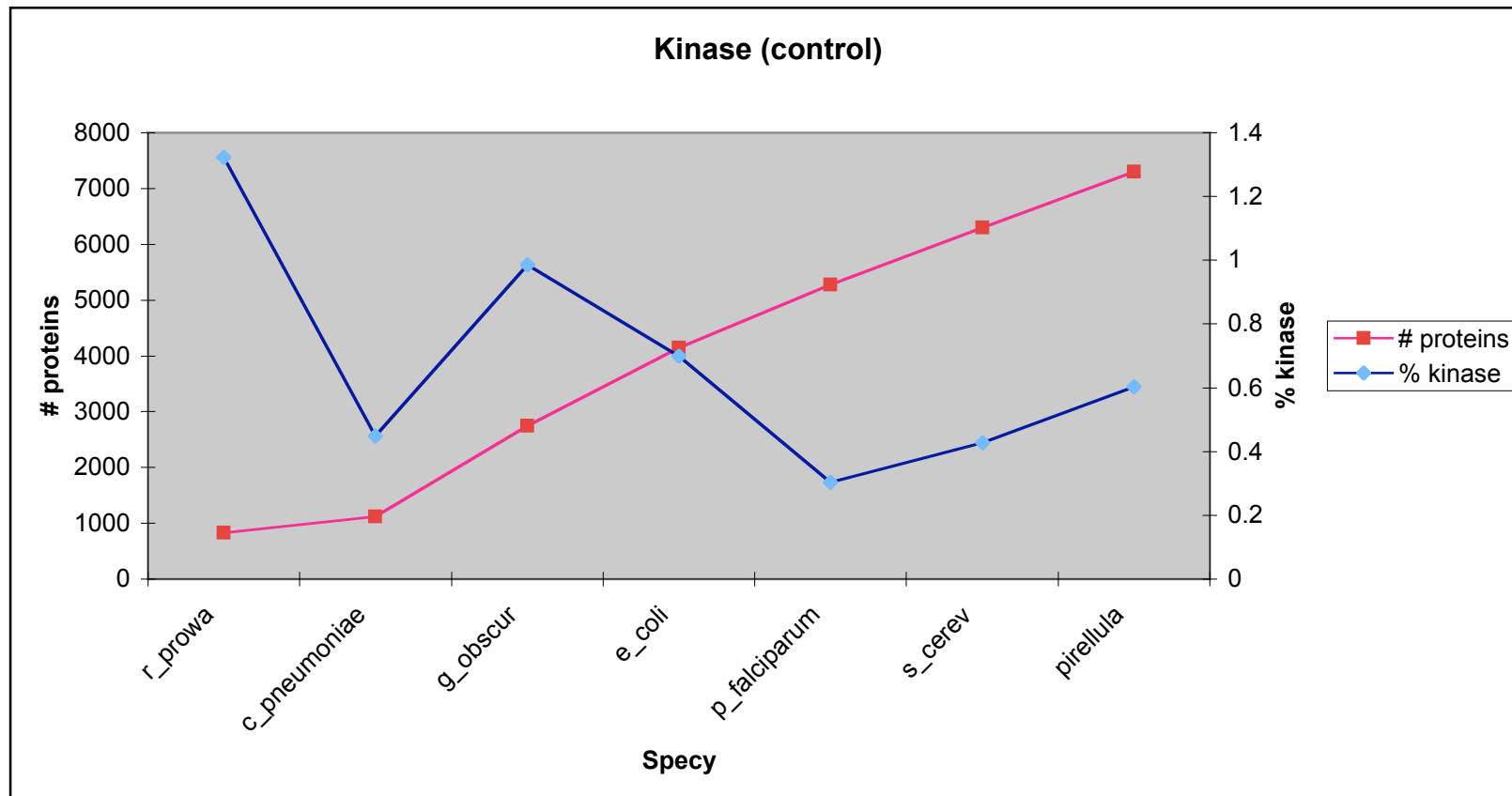
./hhsearch32 -i ${input} -d ${panfs}/scop70_npccv.hhm -o temp.hhs
head -100 temp.hhs >${input}.scop70_npccv.out
$MODPIPEBASE/aux/CopyFile -i ${input}.scop70_npccv.out -o ${panfs}/hhsearch/bacteria/${org}/${input}.scop70_npccv.out
```



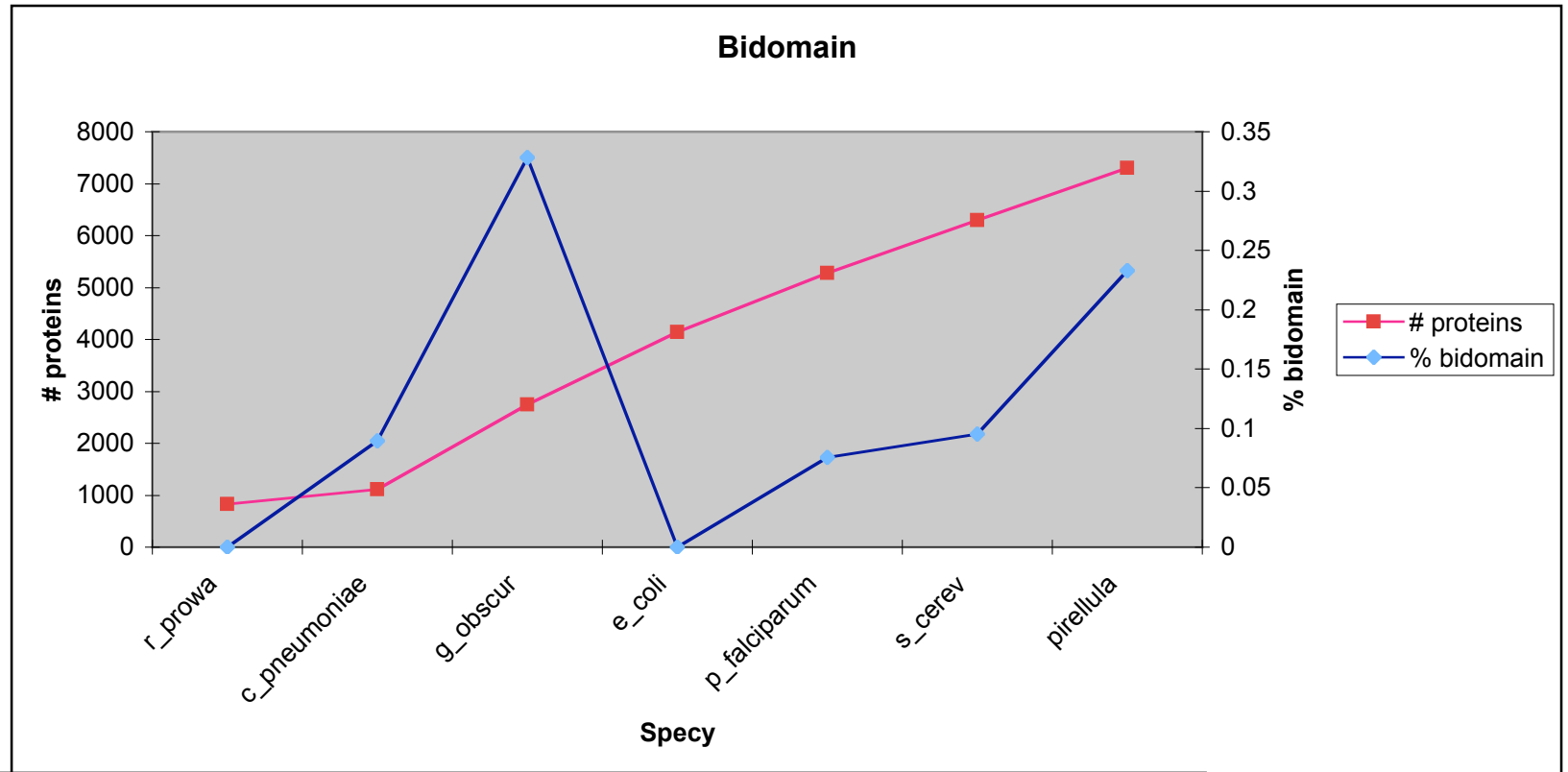
specy	r_prowa	c_pneu moniae	g_obscu r	e_coli	p_falcipa rum	s_cerev	pirellula
# proteins	832	1117	2742	4146	5280	6299	7297
# solenoids	30	35	98	73	412	344	174



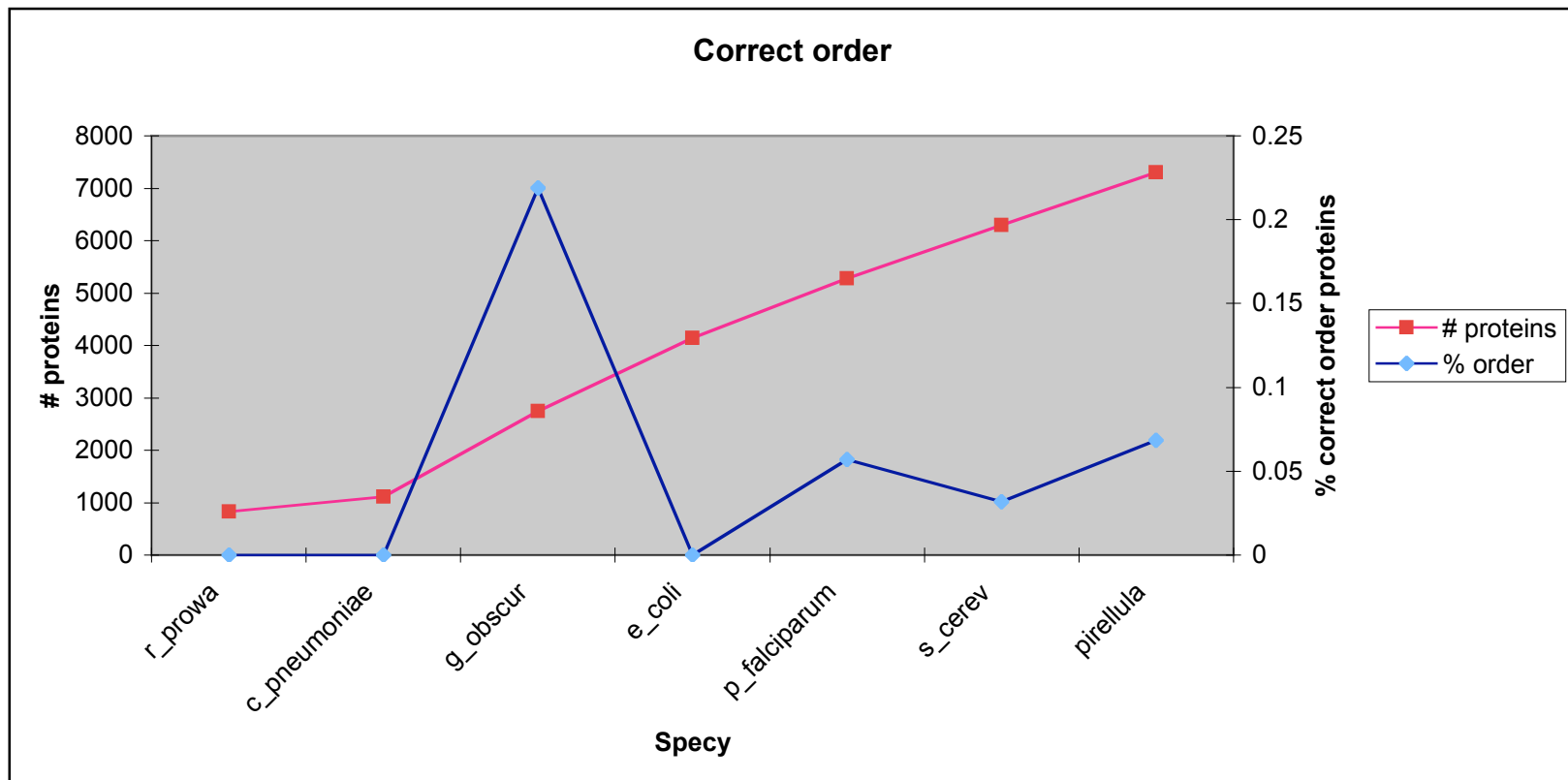
specy	r_prowa	c_pneu moniae	g_obscu r	e_coli	p_falcipa rum	s_cerev	pirellula
# proteins	832	1117	2742	4146	5280	6299	7297
# propeller	9	11	165	52	135	205	220



Specy	r_prowa	c_pneu moniae	g_obscu r	e_coli	p_falcipa rum	s_cerev	pirellula
# proteins	832	1117	2742	4146	5280	6299	7297
# kinase	11	5	27	29	16	27	44

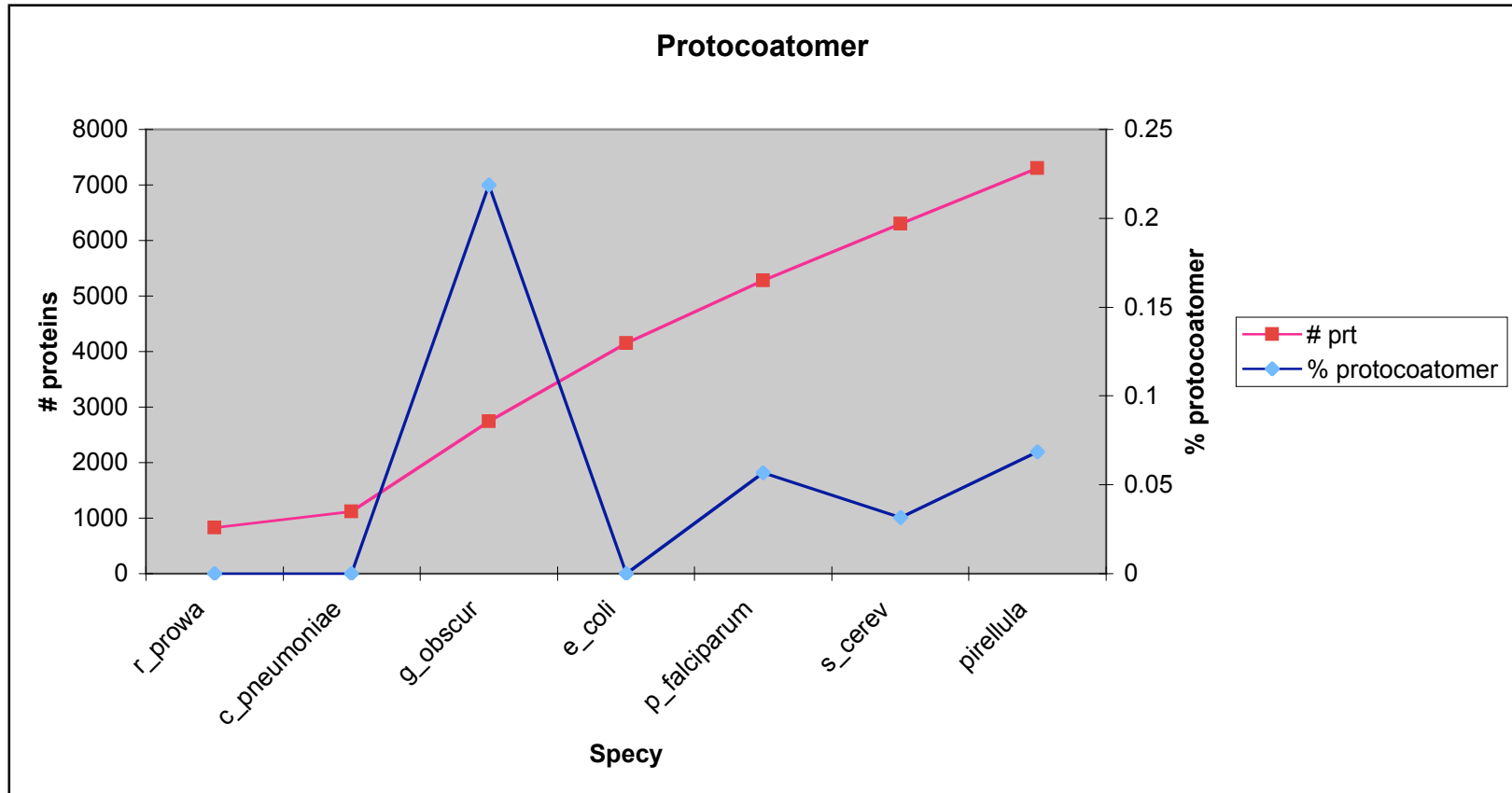


specy	r_prowa	c_pneu moniae	g_obscu r	e_coli	p_falcip arum	s_cerev	pirellula
# proteins	832	1117	2742	4146	5280	6299	7297
# bi domain	0	1	9	0	4	6	17

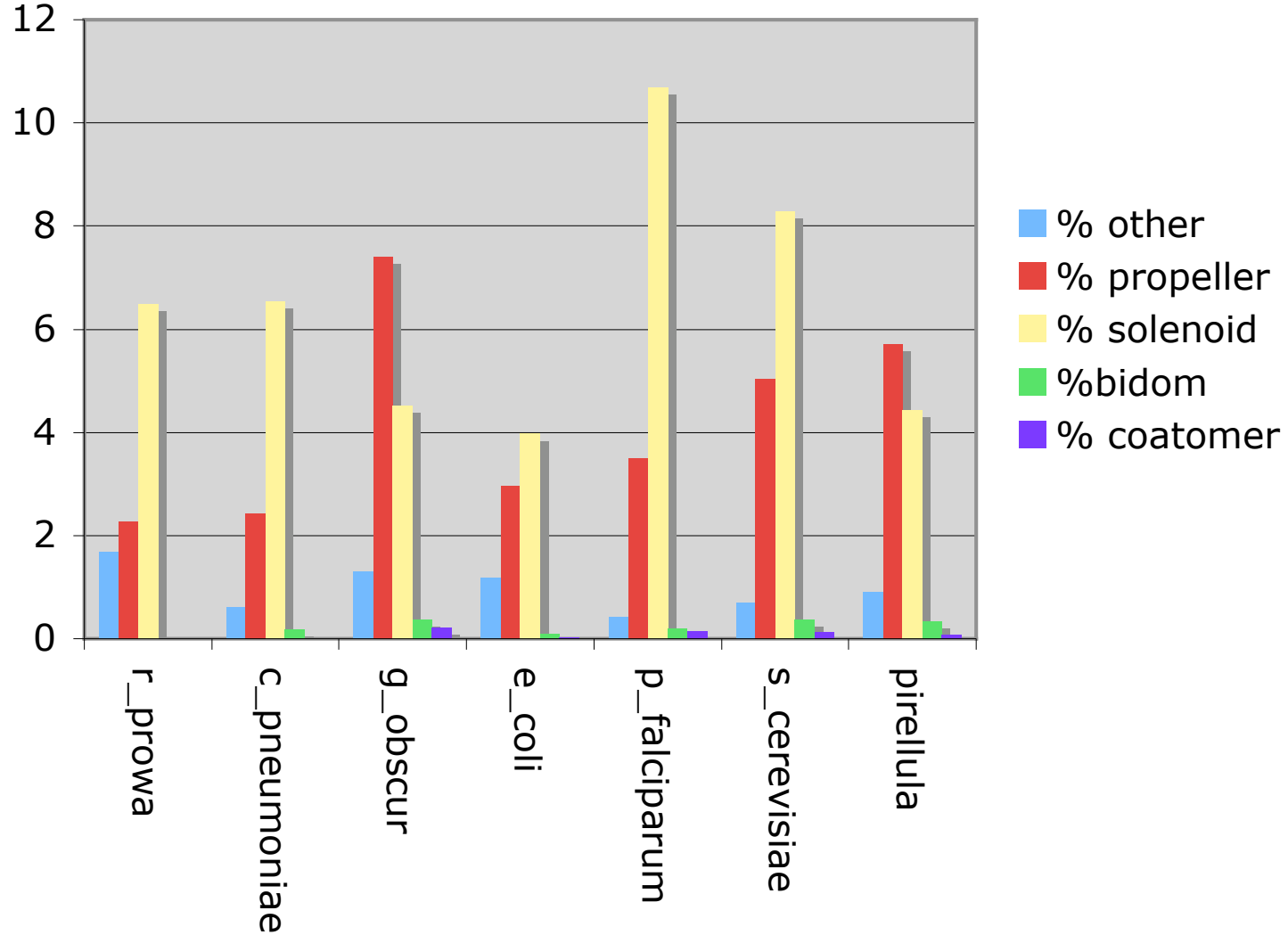


specy	r_prowa	c_pneu moniae	g_obscu r	e_coli	p_falcip arum	s_cerev	pirellula
# proteins	832	1117	2742	4146	5280	6299	7297
# order	0	0	6	0	3	2	5





Specy	r_prowa	c_pneumoniae	g_obscur	e_coli	p_falciparum	s_cerev	pirellula
# proteins	832	1117	2742	4146	5280	6299	7297
# protoatomer	0	0	6	0	3	2	5



proteins	832	1117	2742	4146	5280	6299	7297
other	14	7	36	49	22	44	66
propeller	19	27	203	123	185	317	417
solenoid	54	73	124	165	565	522	324
bidomain	0	2	10	4	10	23	25
coatomer	0	0	6	1	7	8	5

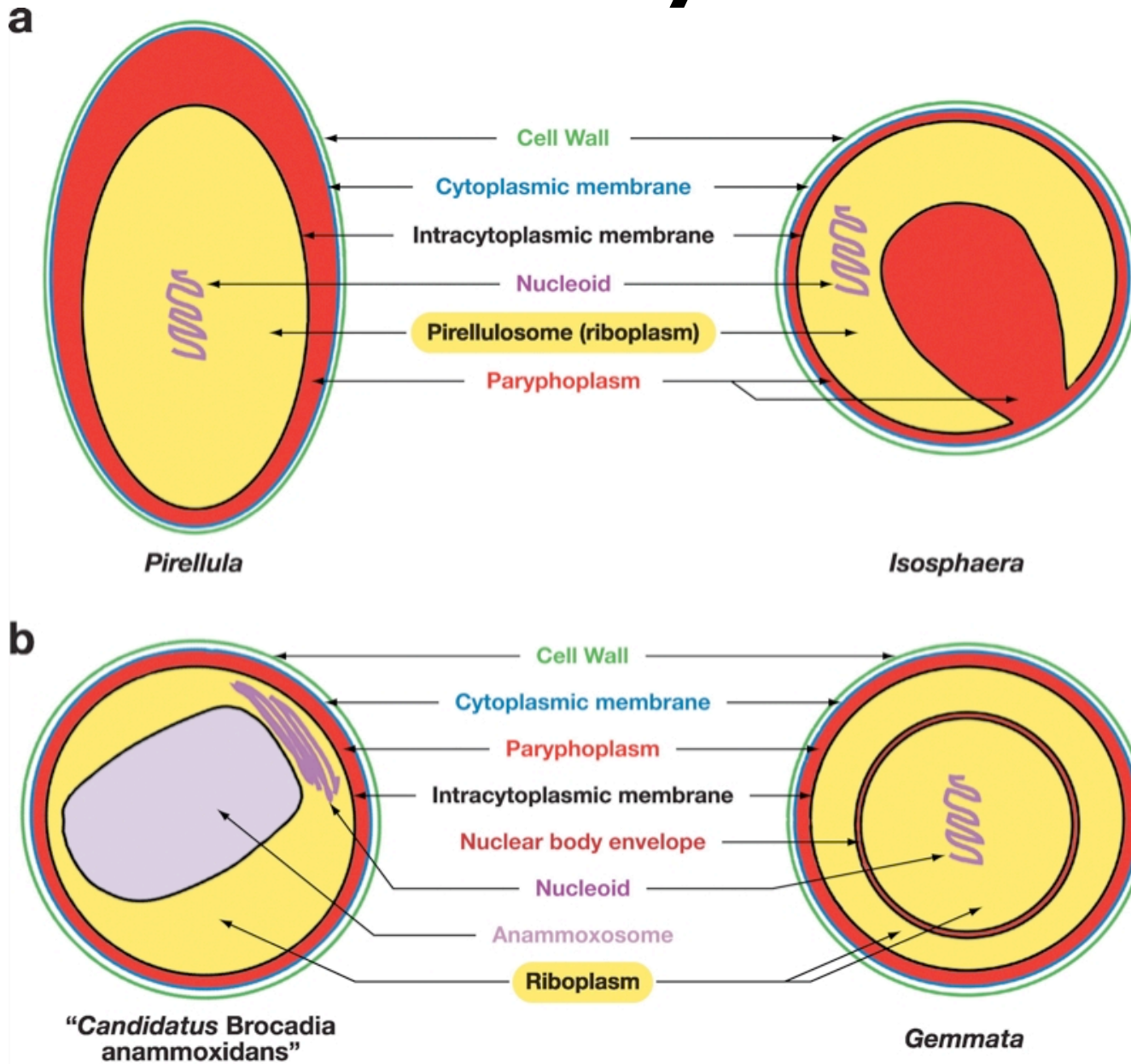
# Archea

Specy	# proteins	# kinase	# propeller	# solenoid	# bidomain	# order	# protoatomer
APER	2696	12	40	32	0	0	0
AFUL	2411	28	28	51	2	1	1
HALO	2607	29	34	25	0	0	0
MJAN	1775	8	11	75	1	0	0
MTHE	1873	9	16	35	0	0	0
PHOR	2063	16	29	56	2	1	1
TVOL	1528	14	20	32	0	0	0
PABY	1767	16	37	47	2	0	0
TACI	1480	17	17	24	0	0	0
SSOL	2997	30	41	92	1	0	0
STOK	2828	23	39	82	0	0	0
PAER	2607	20	36	48	1	0	0
MKAN	1689	10	23	25	0	0	0
MACE	4530	24	119	112	2	0	0
MMAZ	3373	20	63	82	3	1	1
NEQU	565	2	6	15	0	0	0
PTOR	1537	16	20	38	0	0	0
PFUR	2067	19	25	58	0	0	0
MMAR	1724	11	10	52	0	0	0
HMAR	4242	43	67	40	0	0	0
TKOD	2308	25	34	55	1	0	0
SACI	2225	24	32	62	1	0	0
Total	50892						

# Bacteria (I)

Specy	# proteins	# kinase	# propeller	# solenoid	# bidomain	# order	# protoatomer
BBRO	4995	58	68	71	0	0	0
AAEO	1555	20	16	57	0	0	0
CJEI	2139	21	33	26	0	0	0
MGEN	481	7	5	19	0	0	0
SCOE	7812	92	166	105	0	0	0
ACIN	3327	46	42	73	0	0	0
CJEJ	3473	10	44	119	2	0	0
MHYO	693	8	8	30	2	0	0
SENT	22404	117	322	429	1	0	0
AMAR	951	7	9	12	0	0	0
CPER	2725	21	26	86	2	0	0
MLEP	1607	18	18	7	0	0	0
SEPI	2528	24	17	57	1	0	0
ATUM	5301	65	66	66	0	0	0
CPNE	3243	15	33	105	3	0	0
Total	63234						

# Planctomycetes



# Blast hit

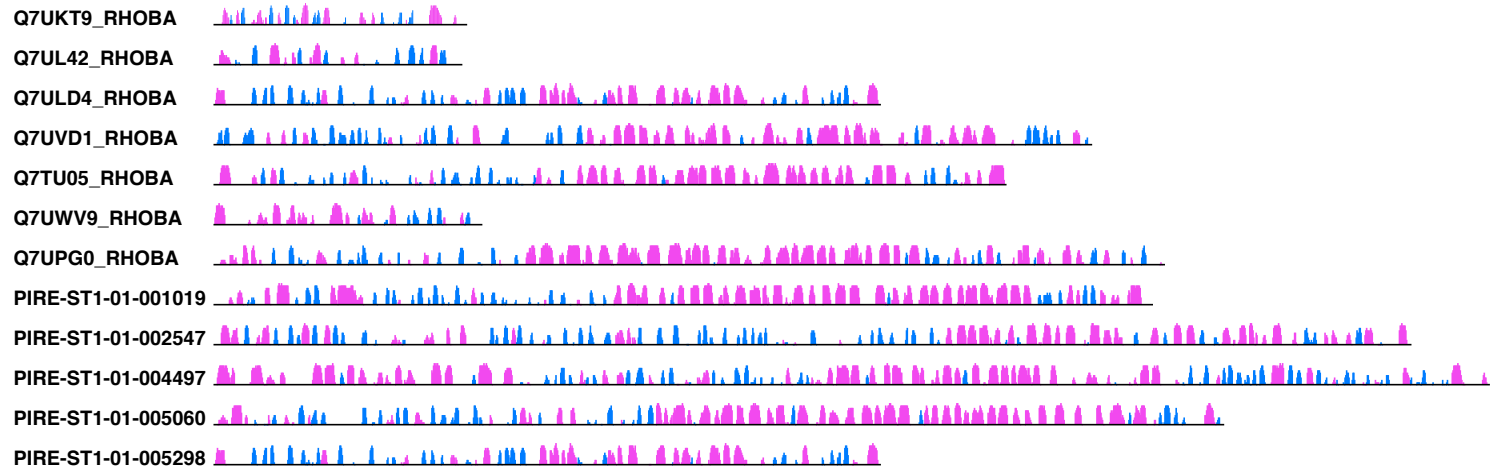
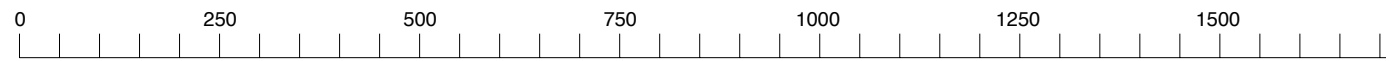
ID/Accession	Protein Name	Organism	Length	SSearch		Blast Search		
				Overlap	%iden	E-Value	Score	Alignment
<input type="checkbox"/> <a href="#">Q7UM20_RHOBA</a> /Q7UM20	RB9120	<a href="#">Rhodopirellula baltica</a>	1263	1263	<u>100</u>	0.0	2538	
<input type="checkbox"/> <a href="#">Q7UYC2_RHOBA</a> /Q7UYC2	RB728	<a href="#">Rhodopirellula baltica</a>	1266	1273	<u>25</u>	5e-72	275	
<input type="checkbox"/> <a href="#">Q7UZE0_RHOBA</a> /Q7UZE0	Similar to L-sorbose dehydrogenase	<a href="#">Rhodopirellula baltica</a>	934	1091	<u>27</u>	1e-61	241	
<input type="checkbox"/> <a href="#">Q7USA4_RHOBA</a> /Q7USA4	RB4621	<a href="#">Rhodopirellula baltica</a>	1497	1125	<u>23</u>	1e-19	102	
<input type="checkbox"/> <a href="#">Q7UT19_RHOBA</a> /Q7UT19	RB4167	<a href="#">Rhodopirellula baltica</a>	979	322	<u>25</u>	1e-15	88.6	
<input type="checkbox"/> <a href="#">Q7UIS5_RHOBA</a> /Q7UIS5	RB12351	<a href="#">Rhodopirellula baltica</a>	1321	437	<u>24</u>	2e-13	81.6	
<input type="checkbox"/> <a href="#">Q7UYL0_RHOBA</a> /Q7UYL0	RB524	<a href="#">Rhodopirellula baltica</a>	1557	1246	<u>20</u>	2e-12	78.2	
<input type="checkbox"/> <a href="#">Q7UV18_RHOBA</a> /Q7UV18	RB2605	<a href="#">Rhodopirellula baltica</a>	1606	1280	<u>20</u>	2e-12	77.8	
<input type="checkbox"/> <a href="#">Q7UPG0_RHOBA</a> /Q7UPG0	Probable L-sorbose dehydrogenase	<a href="#">Rhodopirellula baltica</a>	1189	1092	<u>21</u>	7e-12	75.9	
<input type="checkbox"/> <a href="#">Q7UVD1_RHOBA</a> /Q7UVD1	RB2711	<a href="#">Rhodopirellula baltica</a>	1098	1158	<u>22</u>	9e-12	75.5	
<input type="checkbox"/> <a href="#">Q3XCX7_METFL</a> /Q3XCX7	MflaDRAFT_2654	<a href="#">Methylobacillus flagellatus KT</a>	403	369	<u>26</u>	1e-11	75.1	
<input type="checkbox"/> <a href="#">Q7UMD5_RHOBA</a> /Q7UMD5	RB8892	<a href="#">Rhodopirellula baltica</a>	1360	207	<u>29</u>	4e-11	73.6	

# Modules

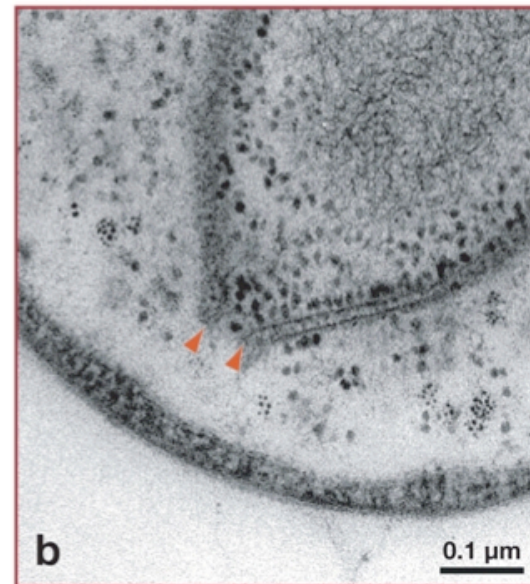
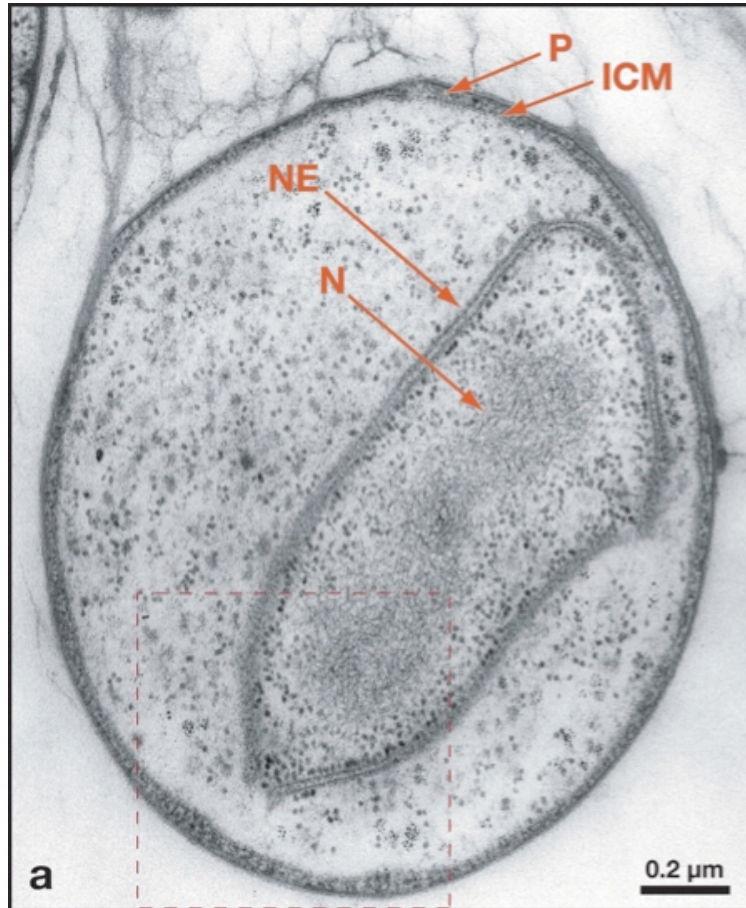
ID/Accession	Protein Name	Organism	Length	SSearch		Blast Search		
				Overlap	%iden	E-Value	Score	Alignment
<input type="checkbox"/> <a href="#">Q7USA4_RHOBA</a> /Q7USA4	RB4621	<a href="#">Rhodopirellula baltica</a>	1497	1497	100	0.0	3051	
<input type="checkbox"/> <a href="#">Q7UQP5_RHOBA</a> /Q7UQP5	RB6185	<a href="#">Rhodopirellula baltica</a>	1074	1043	24	8e-38	162	
<input type="checkbox"/> <a href="#">Q7UKT9_RHOBA</a> /Q7UKT9	RB9945	<a href="#">Rhodopirellula baltica</a>	317	303	36	3e-35	154	
<input type="checkbox"/> <a href="#">Q7UTF2_RHOBA</a> /Q7UTF2	RB3919	<a href="#">Rhodopirellula baltica</a>	1617	1049	24	4e-32	143	
<input type="checkbox"/> <a href="#">Q7UWK3_RHOBA</a> /Q7UWK3	Similar to glucose dehydrogenase-B	<a href="#">Rhodopirellula baltica</a>	1174	1148	24	6e-28	129	
<input type="checkbox"/> <a href="#">Q7UVI8_RHOBA</a> /Q7UVI8	RB2605	<a href="#">Rhodopirellula baltica</a>	1606	729	25	3e-23	114	
<input type="checkbox"/> <a href="#">Q7UYL0_RHOBA</a> /Q7UYL0	RB524	<a href="#">Rhodopirellula baltica</a>	1557	1068	24	2e-21	107	
<input type="checkbox"/> <a href="#">Q7UM20_RHOBA</a> /Q7UM20	RB9120	<a href="#">Rhodopirellula baltica</a>	1263	1125	23	5e-20	103	
<input type="checkbox"/> <a href="#">Q7UL42_RHOBA</a> /Q7UL42	RB9751	<a href="#">Rhodopirellula baltica</a>	311	266	32	2e-18	98.2	
<input type="checkbox"/> <a href="#">Q7ULD4_RHOBA</a> /Q7ULD4	Probable L-sorbose dehydrogenase	<a href="#">Rhodopirellula baltica</a>	834	986	24	2e-18	98.2	
<input type="checkbox"/> <a href="#">Q7UVD1_RHOBA</a> /Q7UVD1	RB2711	<a href="#">Rhodopirellula baltica</a>	1098	1067	23	3e-18	97.1	
<input type="checkbox"/> <a href="#">Q7UPG0_RHOBA</a> /Q7UPG0	Probable L-sorbose dehydrogenase	<a href="#">Rhodopirellula baltica</a>	1189	1075	22	3e-17	94.0	
<input type="checkbox"/> <a href="#">Q7TU05_RHOBA</a> /Q7TU05	RB3322	<a href="#">Rhodopirellula baltica</a>	991	1049	23	2e-16	91.3	





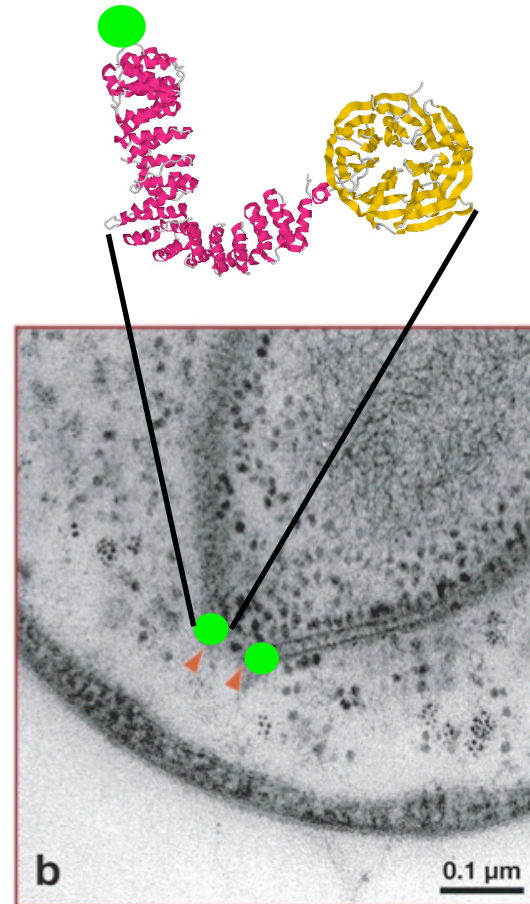
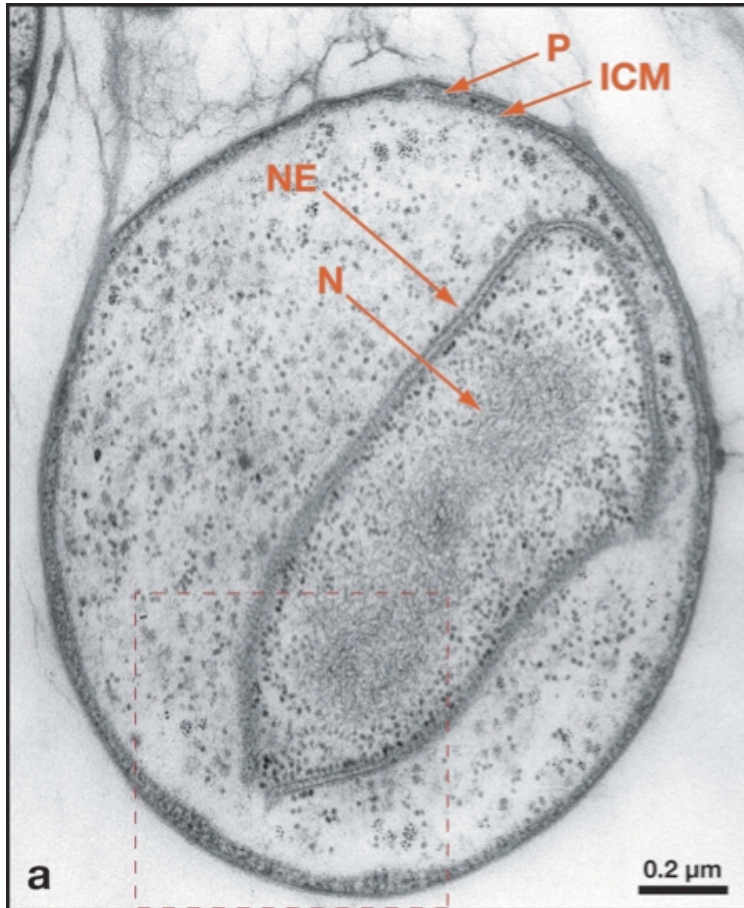


# *Gemmata obscuriglobus*



Fuerst JA. 2005.  
Annu. Rev. Microbiol. 59:299–328

# Simulated data



Fuerst JA. 2005.  
Annu. Rev. Microbiol. 59:299–328