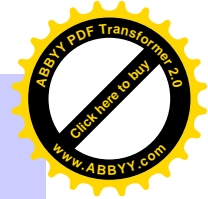
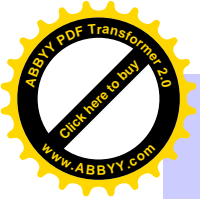


Gos NII Genetika
Moscow, Russia



Hierarchical Classification of Glycoside Hydrolases
Daniil G. Naumoff
Laboratory of Bioinformatics
State Institute for Genetics and Selection of Industrial Microorganisms
Moscow, Russia
<http://bioinform.genetika.ru/members/Naumoff/index.htm>
daniil_naumoff@yahoo.com

Иерархическая классификация гликозид гидролаз
Даниил Геннадиевич Наумов
Лаборатория биоинформатики
Государственный Научно-Исследовательский Институт Генетики и Селекции
Промышленных Микроорганизмов, Москва



Enzyme Classification (IUBMB)

EC 1 – Oxidoreductases

EC 2 – Transferases

EC 2.4 – Glycosyltransferases (including transglycosidases)

EC 2.4.1 – **Hexosyltransferases**

EC 2.4.1.10 – **Levansucrase (β -fructosyltransferase)**

EC 2.4.1.67 – **Galactinol-raffinose galactosyltransferase**

EC 2.4.2 – Pentosyltransferases

EC 3 – Hydrolases

EC 3.2 – Glycosylases (glycoside hydrolases)

EC 3.2.1 – **O- and S-Glycosylases**

EC 3.2.1.20 – **Maltase (α -glucosidase)**

EC 3.2.1.22 – **Melibiase (α -galactosidase)**

EC 3.2.1.26 – **Sucrase (invertase, β -fructofuranosidase)**

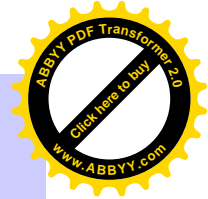
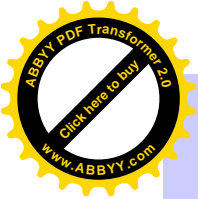
EC 3.2.2 – N-Glycosylases

EC 4 – Lyases

EC 5 – Isomerases

EC 6 – Ligases

Glycoside Hydrolases = EC 3.2.1.X; X=1...>200



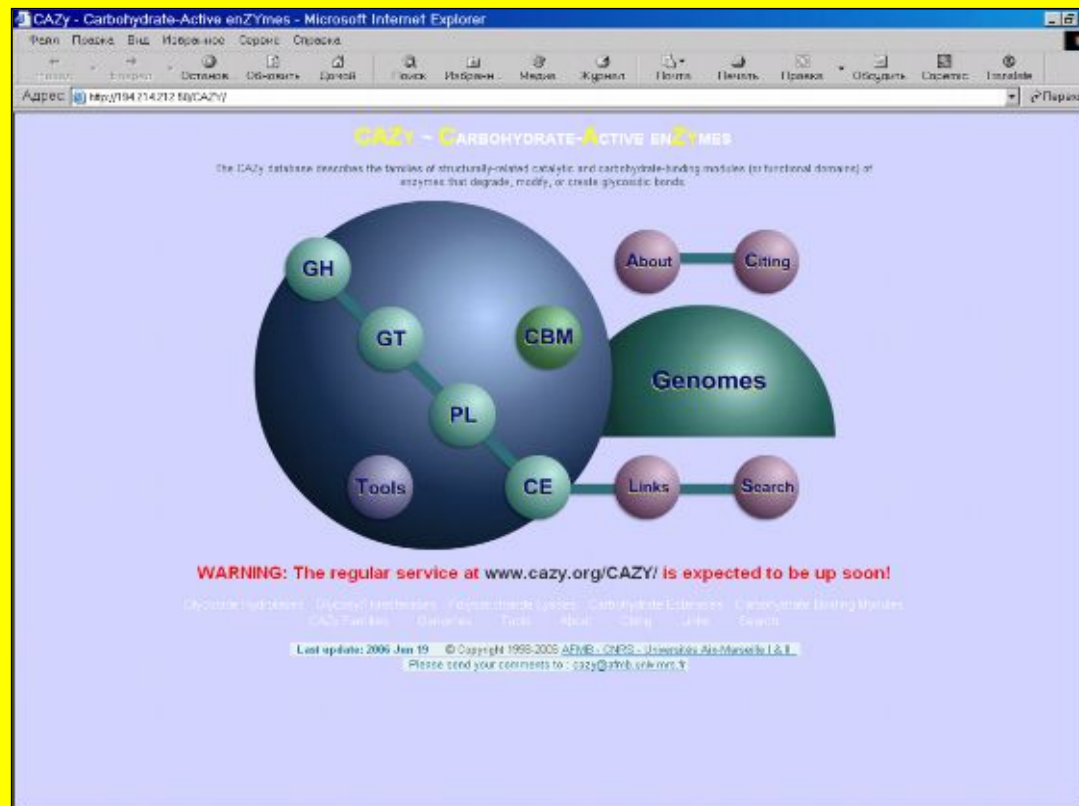
Progress in the Sequence Based Classification of Glycoside Hydrolases and Their Homologues

Number of proteins	Number of families	Reference
300	36	Henrissat. <i>Biochem. J.</i> 1991, 280:309-316
480	45	Henrissat & Bairoch. <i>Biochem. J.</i> 1993, 293:781-788
950	57 (17 => 5 clans)	Henrissat & Bairoch. <i>Biochem. J.</i> 1996, 316:695-696
...	62 (19 => 5 clans)	Henrissat & Davies. <i>Curr. Opin. Struct. Biol.</i> 1997, 7:637-644
2.200	70 (...)	Coutinho & Henrissat. <i>CAZy server.</i> 1998
>20.000	101 (46 => 14 clans)	Coutinho & Henrissat. <i>CAZy server.</i> 2005
~30.000	106 (46 => 14 clans)	Coutinho & Henrissat. <i>CAZy server.</i> 2006
>30.000	110 (46 => 14 clans)	Coutinho & Henrissat. <i>CAZy server.</i> 2007

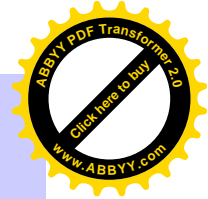
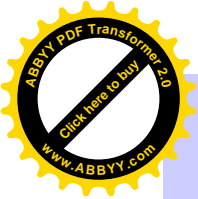
CAZy = the Carbohydrate-Active enZYmes server (<http://www.cazy.org/CAZY>)

Sequence Based Classification of Glycoside Hydrolases

the Carbohydrate-Active Enzymes (CAZy) server (www.cazy.org/CAZY/)

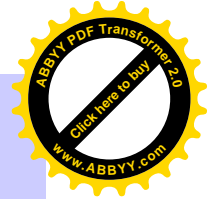
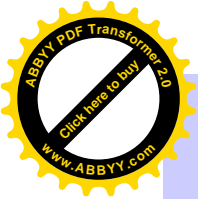


- 30.000 protein sequences => 105 families (GH1-GH110, except GH21, GH40, GH41, GH60, and GH69)
- 46 glycoside hydrolase families => 14 clans (GH-A–GH-N)



Clans of Glycoside Hydrolases

Clan	Families (GH)	Optical Configuration	Tertiary Structure
GH-A	1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 50, 51, 53, 59, 72, 79, 86	retention (equatorial orientation)	(β/α)₈ -barrel
GH-B	7, 16	retention (equatorial orientation)	β -jelly roll
GH-C	11, 12	retention (equatorial orientation)	β -jelly roll
GH-D	27, 36	retention (axial orientation)	(β/α)₈ -barrel
GH-E	33, 34, 83	retention (equatorial orientation)	6-fold β -propeller
GH-F	43, 62	inversion (equatorial orientation)	5-fold β -propeller
GH-G	37, 63	inversion (axial orientation)	Not known
GH-H	13, 70, 77	retention (axial orientation)	(β/α)₈ -barrel
GH-I	24, 46, 80	inversion (equatorial orientation)	$\alpha+\beta$
GH-J	32, 68	retention (β -furanoside)	5-fold β -propeller
GH-K	18, 20	retention (equatorial orientation)	(β/α)₈ -barrel
GH-L	15, 65	inversion (axial orientation)	(α/α) ₆
GH-M	8, 48	inversion (equatorial orientation)	(α/α) ₆
GH-N	28, 49	inversion (axial orientation)	(β) ₃ -solenoid



EC 3.2.1.22

Common name: α -galactosidase

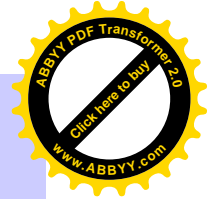
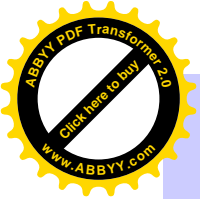
Reaction: Hydrolysis of terminal, non-reducing α -D-galactose residues in α -D-galactosides, including galactose oligosaccharides, galactomannans, and galactolipids

Molecular mechanism of hydrolyzing reaction: double displacement with overall retention of the anomeric configuration of the axial glycosidic bond

Other names: melibiase; α -D-galactosidase; α -galactosidase A; α -D-galactoside galactohydrolase; α -galactoside galactohydrolase

Systematic name: α -D-galactoside galactohydrolase

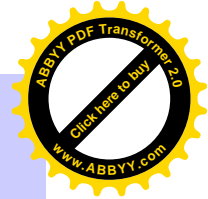
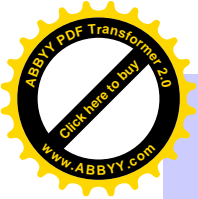
Comments: Also hydrolyses α -D-fucosides



Glycoside hydrolase families, which include α -galactosidases

- GH4** – α -galactosidase from *Escherichia coli* and bifunctional enzymes with α -galactosidase and α -glucosidase (EC 3.2.1.20) activities from *Thermotoga maritime* and *Thermotoga neapolitana*
+ 6-phospho- α -glucosidases (EC 3.2.1.122) from Eubacteria
+ 6-phospho- β -glucosidases (EC 3.2.1.86) from Eubacteria
- GH27** – mainly α -galactosidases from Eukaryota (animals, plants, fungi, etc.) and some Eubacteria
+ α -N-acetylgalactosaminidases (EC 3.2.1.49) from Eukaryota (animals and fungi)
+ isomalto-dextranase (EC 3.2.1.94) from *Arthrobacter globiformis*
- GH36** – mainly α -galactosidases from Eubacteria and Eukaryota (fungi and plants)
+ α -galactosyltransferases (EC 2.4.1.67 and EC 2.4.1.82) from plants
+ α -N-acetylgalactosaminidase from *Clostridium perfringens*
+ ORFs from *Sulfolobus solfataricus* and *Sulfolobus tokodaii*
- GH57** – α -galactosidases from *Pyrococcus furiosus* and *Thermococcus alcaliphilus*
+ α -amylases (EC 3.2.1.1), amylopullulanase (EC 3.2.1.41), and
4- α -glucanotransferases (EC 2.4.1.x) from Eubacteria and Archaea
- GH110** – α -galactosidase from *Bacteroides fragilis*

clan
GH-D



Family GH27

Subfamily 27a:

- plant α -galactosidases: *Arabidopsis*, *Carica*, *Coffea*, *Cyamopsis*, *Glycine*, *Helianthus*, *Hordeum*, *Lycopersicon*, *Oryza*, *Petunia*, *Phaseolus*, and *Senna*
- α -galactosyltransferase from *Ajuga reptans*
- animal α -galactosidases and α -N-acetylgalactosaminidases: *Anopheles*, *Apis*, *Ateles*, *Brugia*, *Caenorhabditis*, *Ciona*, *Danio*, *Drosophila*, *Gallus*, *Homo*, *Mus*, *Rattus*, *Takifugu*, and *Tetraodon*
- fungal and yeast α -galactosidases: *Aspergillus*, *Ganoderma*, *Gibberella*, *Magnaporthe*, *Mortierella*, *Penicillium*, *Phanerochaete*, *Saccharomyces*, *Schizosaccharomyces*, *Thermomyces*, *Torulaspora*, *Trichoderma*, *Ustilago*, and *Zygosaccharomyces*
- α -N-acetylgalactosaminidase from *Acremonium* sp.
- ORFs from *Dictyostelium discoideum* and *Toxoplasma gondii*
- bacterial α -galactosidases: *Cellvibrio mixtus*, *Clostridium josui*, *Pseudomonas fluorescens*, *Saccharopolyspora erythraea*, and *Streptomyces coelicolor*
- bacterial ORFs: *Bacteroides*, *Fibrobacter*, *Microbulbifer*, *Porphyromonas*, and *Streptomyces*

Subfamily 27b:

- plant ORFs: *Arabidopsis thaliana* and *Oryza sativa*
- bacterial ORFs: *Bacillus halodurans*, *Bifidobacterium longum*, *Kineococcus radiotolerans*, and *Ruminococcus albus*

Subfamily 27c:

- α -galactosidase from *Trichoderma reesei*
- ORF from *Aspergillus nidulans*

Unclassified proteins

- isomalto-dextranase from *Arthrobacter globiformis*
- bacterial ORFs: *Bacteroides thetaiotaomicron* and *Streptomyces avermitilis*

Tertiary Structure of GH27 Glycosidases (*Trichoderma reesei* α -galactosidase)



GH27N

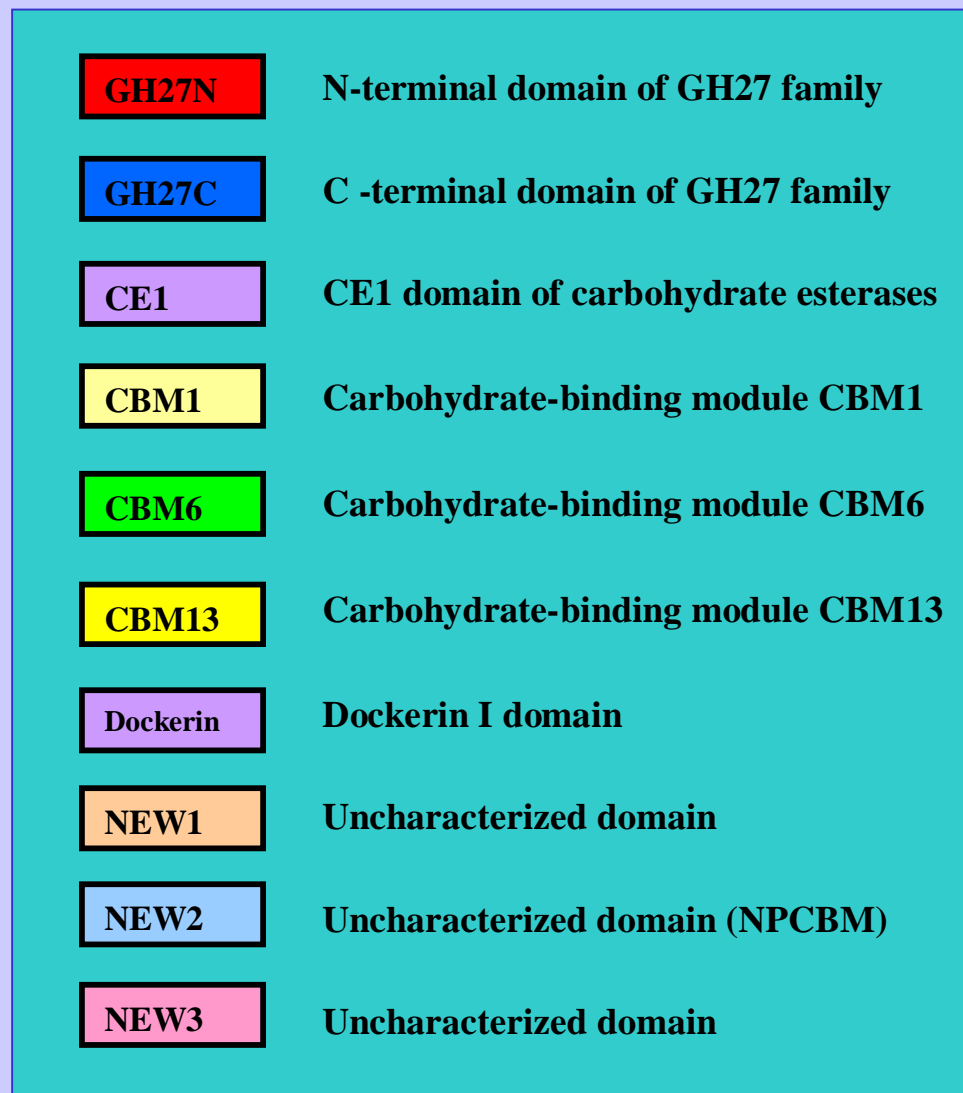
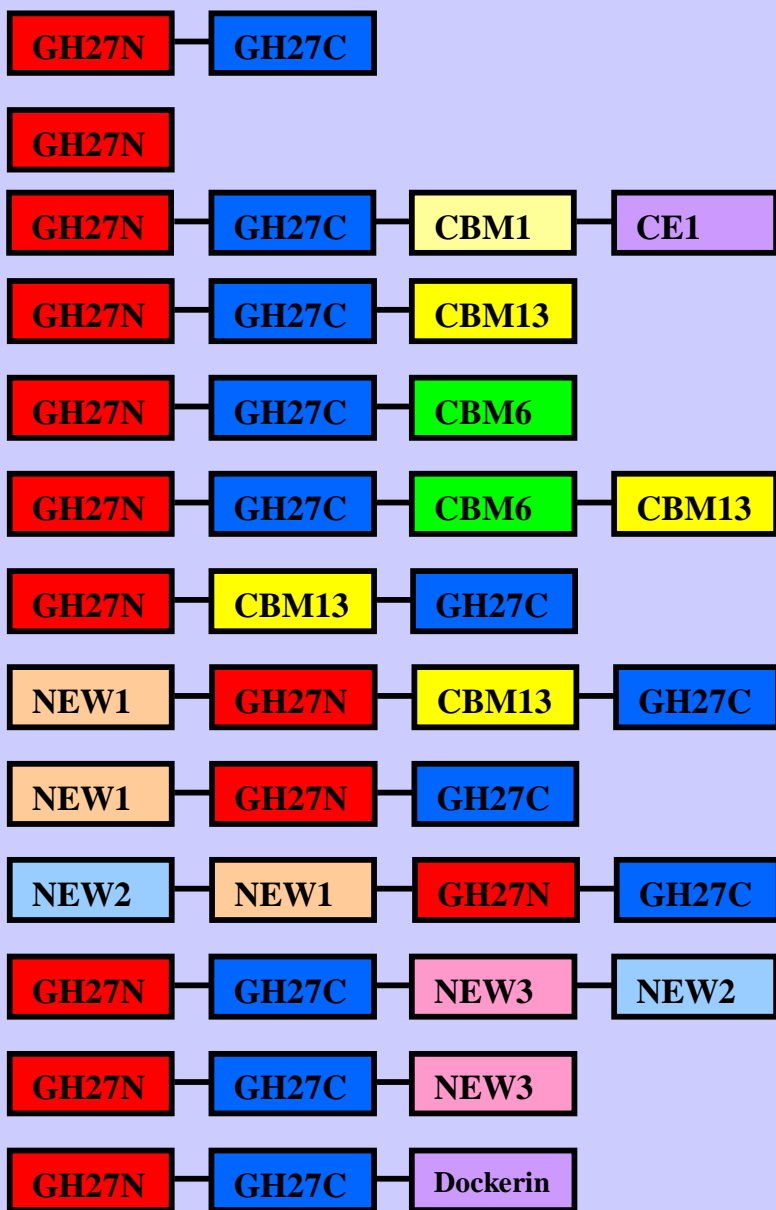
N-terminal catalytic
(β/α)₈-domain
(TIM barrel-type structure)

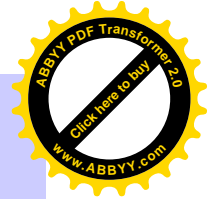
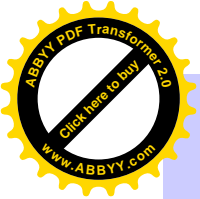
GH27C

C-terminal domain
(β -sandwich fold,
Greek key motif)

Domain structure of proteins of the GH27 family

according to Naumoff D.G. Phylogenetic analysis of α -galactosidases of the GH27 family. Molecular Biology (Engl Transl), 2004, 38(3):388-399.
PDF: <http://bioinform.genetika.ru/members/Naumoff/MB2004E.pdf>





Family GH36

Subfamily GH36A:

- bacterial α -galactosidases: *Azotobacter*, *Bacillus*, *Geobacillus*, *Bifidobacterium*, *Carnobacterium*, *Clostridium*, *Enterococcus*, *Erwinia*, *Escherichia*, *Geobacillus*, *Klebsiella*, *Lactobacillus*, *Lactococcus*, *Leuconostoc*, *Novosphingobium*, *Oenococcus*, *Pediococcus*, *Ruminococcus*, *Streptococcus*, *Streptomyces*, *Thermoanaerobacter*, *Treponema*, *Vibrio*, and *Yersinia*
- fungal α -galactosidases: *Absidia*, *Aspergillus*, *Gibberella*, *Penicillium*, and *Trichoderma*

Subfamily GH36B:

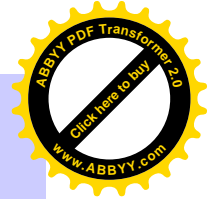
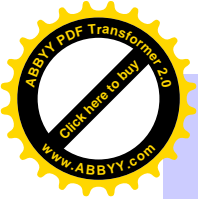
- bacterial α -galactosidases: *Burkholderia*, *Leptospira*, *Streptomyces*, *Thermotoga*, *Thermus*, and *Vibrio*

Subfamily GH36C:

- plant alkaline α -galactosidases or seed imbibition proteins: *Arabidopsis*, *Brassica*, *Cicer*, *Cucumis*, *Hordeum*, *Lycopersicon*, *Malus*, *Oryza*, and *Persea*
- plant α -galactosyltransferases: *Alonsoa*, *Arabidopsis*, *Cucumis*, *Medicago*, *Oryza*, *Pisum*, *Stachys*, and *Vigna*
- fungal ORFs: *Aspergillus*, *Coccidioides*, *Cryptococcus*, *Gibberella*, *Magnaporthe*, *Neurospora*, and *Ustilago*
- ORF from *Toxoplasma gondii*
- α -galactosidase from *Bifidobacterium breve*
- bacterial ORFs: *Bifidobacterium longum*, *Bacteroides fragilis*, and *Bacteroides thetaiotaomicron*
- ORFs from *Sulfolobus solfataricus* and *Sulfolobus tokodaii*

Subfamily GH36D:

- α -N-acetylgalactosaminidase from *Clostridium perfringens*
- bacterial ORFs: *Aeromonas*, *Clostridium*, *Escherichia*, *Photobacterium*, and *Streptococcus*



Four conserved regions in proteins of the α -galactosidase superfamily, containing Asp residues

MEL1_YEAST	65	GYKYIILDDCWS	141	NRVDYLKYDNCY	204	WRMSGDV	258	WNDLDNL	
AGAL_PHAVU	106	GYQYINIDDCWG	184	WGIDYLYKYNCE	242	WRTTGDI	274	WNDPDML	
NAGA_HUMAN	71	GYTYLNIDDCWI	148	WKVDMLKLDGCF	212	WRNYDDI	248	WNDPDML	
AGL1_BIFLO	48	GWDTLVIDIDWY	178	WGLDFLKVDDMQ	240	WRISDDL	272	WADADMV	
AGL3_HYPJE	234	GYDLCSLDSGWQ	339	WGVDMMLKLDFLT	407	MRTDQDL	451	YPDMDAL	
IMD_ARTGO	100	GYDIACTD-GWI	219	LGVPYLRIDFLS	290	VRINADA	338	ILDGDFM	
AGAL_LACPL	367	GIEMFVLDDGWF	478	VPIDYIKWDMNR	551	QSWPSDN	584	GTSPDEL	
AGL2_HYPJE	375	GIKLFVLDDGWF	496	ASISYVKWDMNR	561	HIWTSDD	594	SAVPNGQ	
AGL7_ASPFU	413	GAGYFVIDAGWY	524	YGVGYFKFDYNI	595	LQSSSDQ	626	WAYPQPA	
AGAL_THEMA	213	PFEVFQIDDAYE	319	MGYRYFKIDFLF	382	MRIGPDT	423	LNDPDCL	
AGAL_VIBPA	227	DLEWVLLDDGYQ	341	WGVLEFKLDANY	404	MRVSDDV	436	QIDPDCA	
AGL3_STRCO	329	GLKWAVLDDGWQ	438	WGYEGLKIDGQH	512	QYPSDDP	540	SYSGDHV	
AGAL_SULSO	256	RLNWVIIDDGWQ	360	RDFDLVKVDNQW	421	MRNSIDY	454	YPDYDMF	
AGAL_BIFBR	339	PVSWVLIDDGWS	458	AGVDFVKVDSQS	523	TRTSDDF	556	HCDWDMF	
GALT_VIGAN	251	APRFVVIDDGWQ	476	TGVTGVKIDVIH	544	GRVGDDF	585	QPDWDMF	
NAGA_CLOPE	253	TLDAFVVDGWA	355	YDISYWKIDGML	429	IQTSQDV	...		
ORF1_CLOPE	515	PIDSYVVDGWH	634	FDIDYWKLDGFA	715	IQNSQDT	...		
ORF1_ECOLI	297	ALDAFLLDDGWD	393	EHITSFKLDGMG	458	WRQGDDI	...		
ORF2_CLOPE	112	PKGIIMIDDGWS	220	YGVDFGFKFDAGD	286	HSWEYNG	344	ALMPMMQ	
XYLQ_LACPE	297	PLDVHFHDFCFWQ	406	MGVDSFKTDFGE	474	IQYTGAA	534	LLSSHRS	
XYLS_SULSO	250	PLDVIVLDWRYW	345	LGIDAYWLDASE	423	ISWSGDV	483	TFCPILR	
LYAG_HUMAN	397	PLDVQWNLDDYM	510	VPFDGMWIDMNE	611	GHWTGDV	666	AFYPFMR	

GH27

GH36A

GH36B

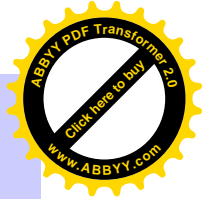
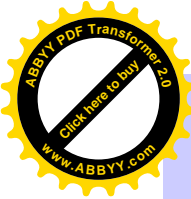
GH36C

GH36D

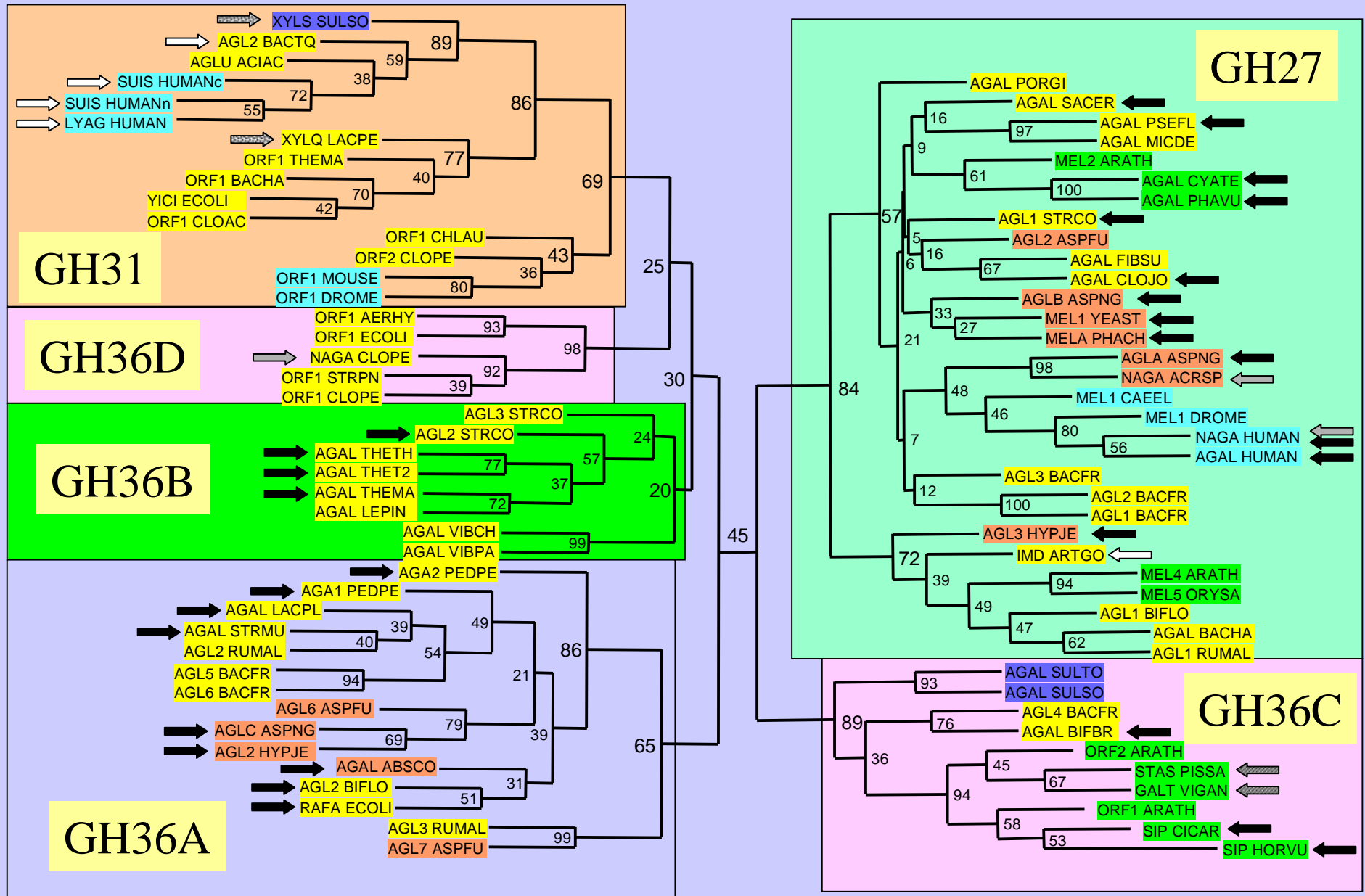
GH31

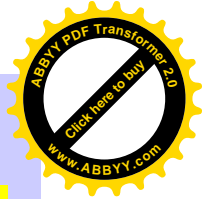
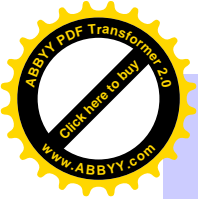
Nucleophile ↑

↑ Donor H⁺



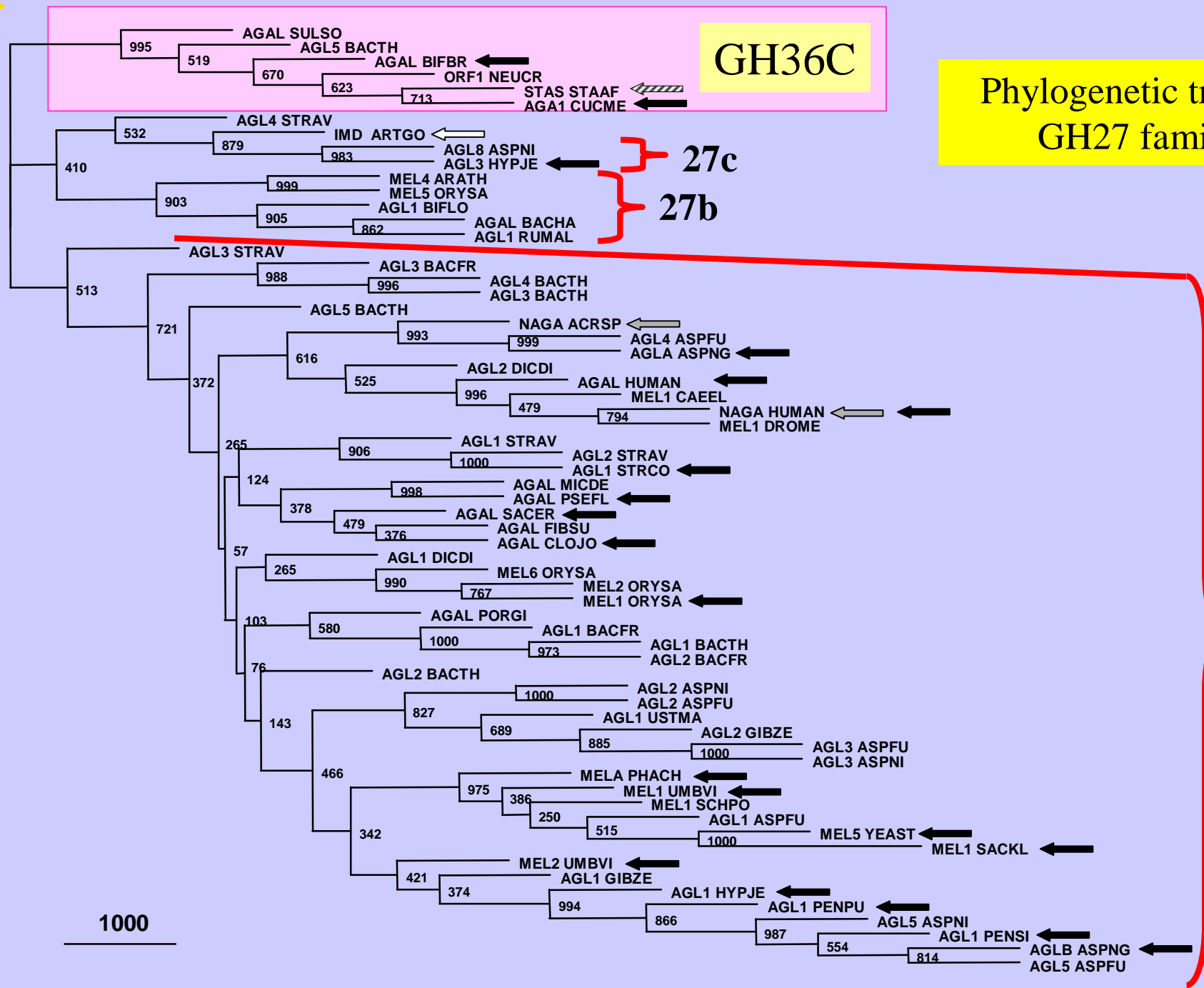
The neighbor-joining phylogenetic tree of the α -galactosidase superfamily

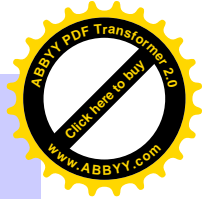
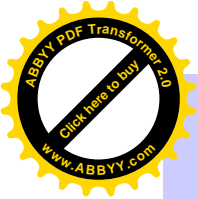




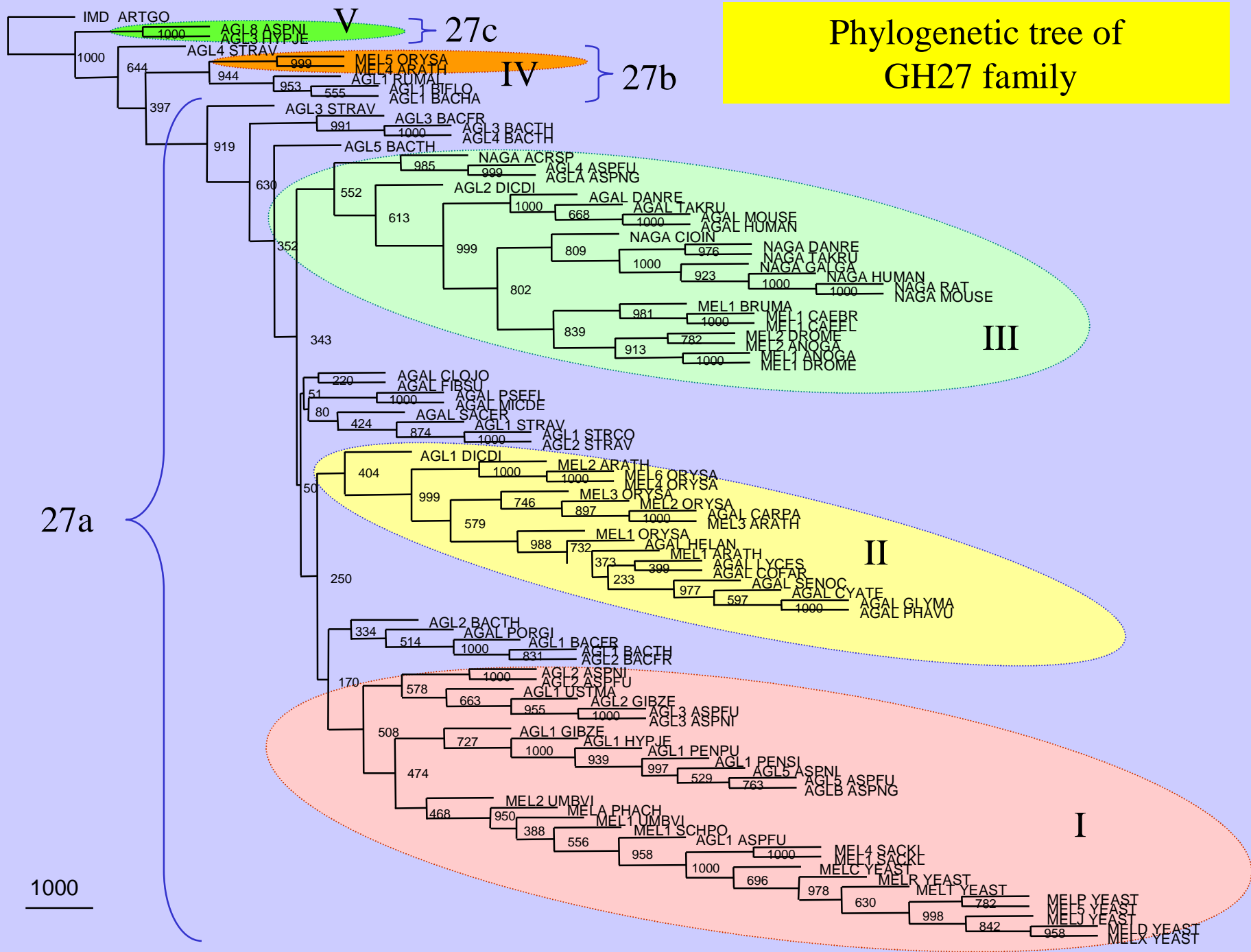
Phylogenetic tree of
GH27 family

GH36C

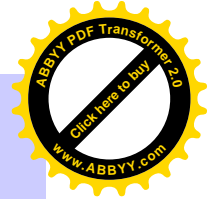
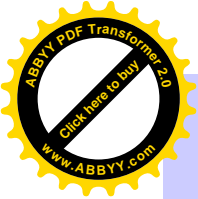




Phylogenetic tree of GH27 family

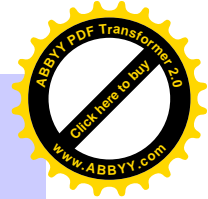
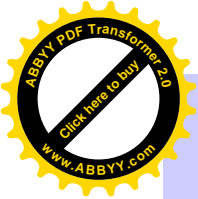


1000

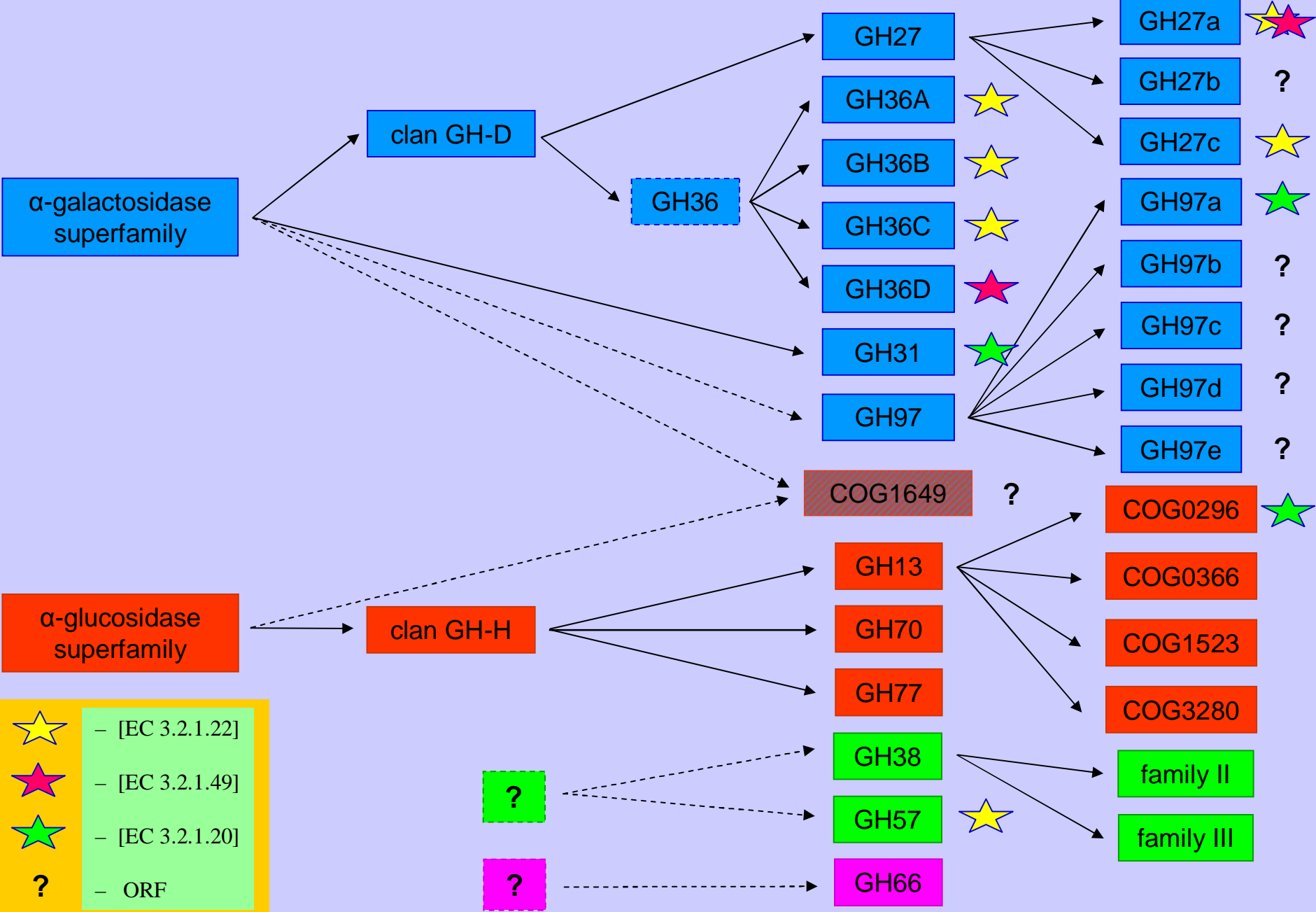


Families of the α -galactosidase superfamily and family GH97

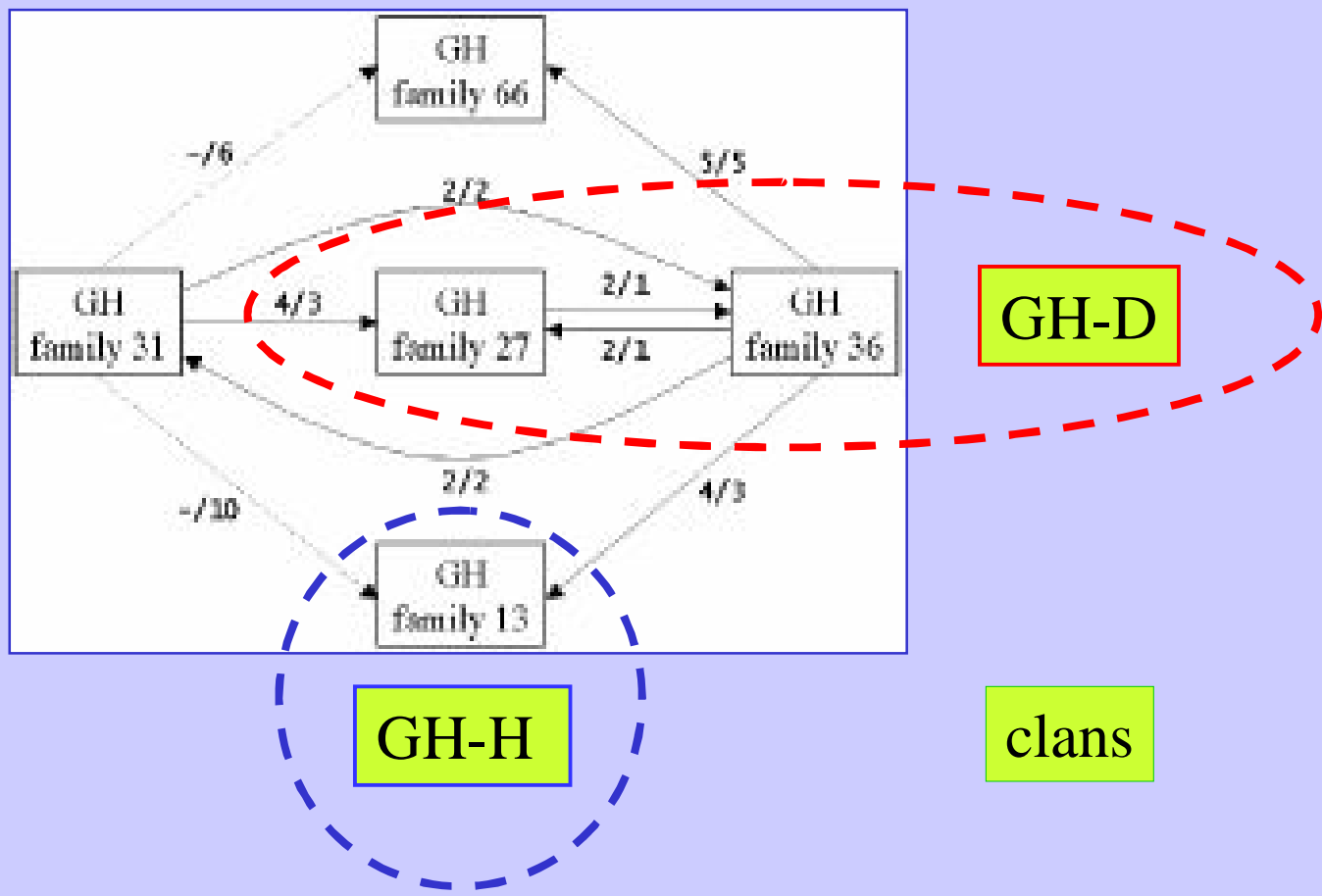
Family	GH27	GH31	GH36A	GH36B	GH36C	GH36D	GH97
Clan	GH-D	None	GH-D	GH-D	GH-D	GH-D	None
COG/KOG	KOG2366	COG1501 KOG1065	COG3345	COG3345	None	None	None
Known enzymatic activities	EC 2.4.1.x EC 3.2.1.22 EC 3.2.1.49 EC 3.2.1.94	EC 2.4.1.x EC 3.2.1.3 EC 3.2.1.10 EC 3.2.1.20 EC 3.2.1.48 EC 4.2.2.13	EC 3.2.1.22	EC 3.2.1.22	EC 2.4.1.67 EC 2.4.1.82 EC 3.2.1.22	EC 3.2.1.49	EC 3.2.1.20
Molecular mechanism	Retaining	Retaining	Retaining	Not known	Not known	Not known	Not known
Origin	Eukaryota: Alveolata Fungi Metazoa Mycetozoa Viridiplantae Eubacteria: Acidobacteria Actinobacteria Bacteroidetes Fibrobacteres Firmicutes Proteobacteria	Eukaryota: Alveolata Entamoebidae Euglenozoa Fungi Metazoa Mycetozoa Rhodophyta Viridiplantae Eubacteria: Acidobacteria Actinobacteria Bacteroidetes Cyanobacteria Firmicutes Proteobacteria Spirochaetes Thermotogales Thermus Archaea: Crenarchaeota Euryarchaeota	Eukaryota: Fungi Eubacteria: Actinobacteria Bacteroidetes Firmicutes Proteobacteria Spirochaetes	Eubacteria: Acidobacteria Actinobacteria Deinococcus Proteobacteria Spirochaetes Thermotogales Thermus	Eukaryota: Alveolata Fungi Viridiplantae Eubacteria: Actinobacteria Bacteroidetes Archaea: Crenarchaeota	Eubacteria: Firmicutes Proteobacteria	Eukaryota: Metazoa (?) Eubacteria: Acidobacteria Actinobacteria Bacteroidetes Firmicutes Planctomycetes Proteobacteria Archaea: Euryarchaeota

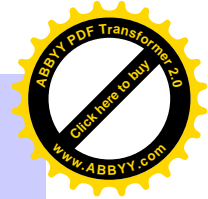
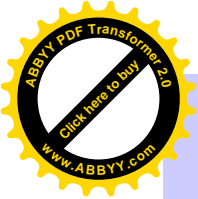


(β/α)₈-barrel-type retaining α -D-glycopyranosidase families

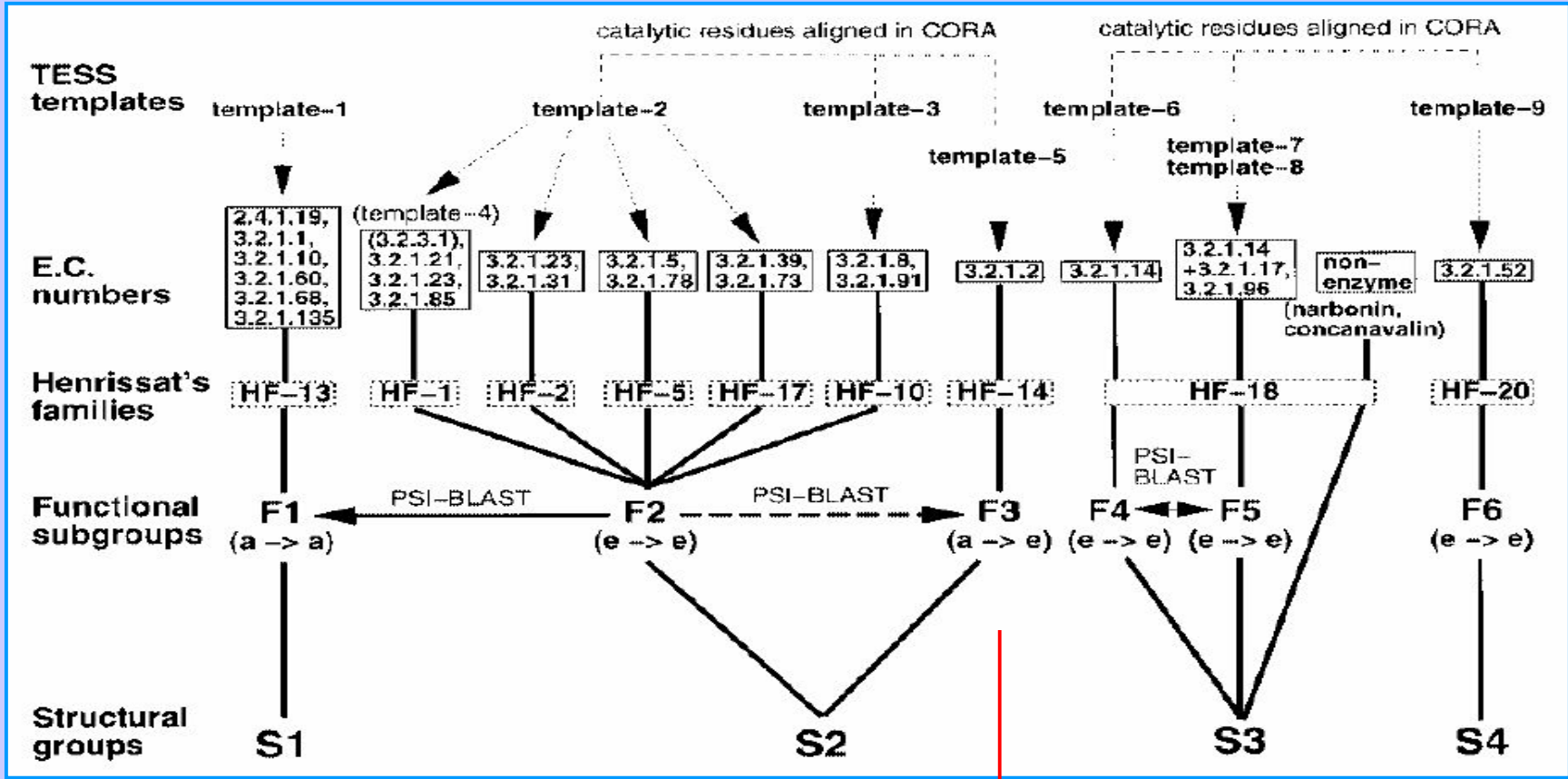


Rigden DJ. Iterative database searches demonstrate that glycoside hydrolase families 27, 31, 36, and 66 share a common evolutionary origin with family 13. *FEBS Lett.* 2002, 523(1-3):17-22.

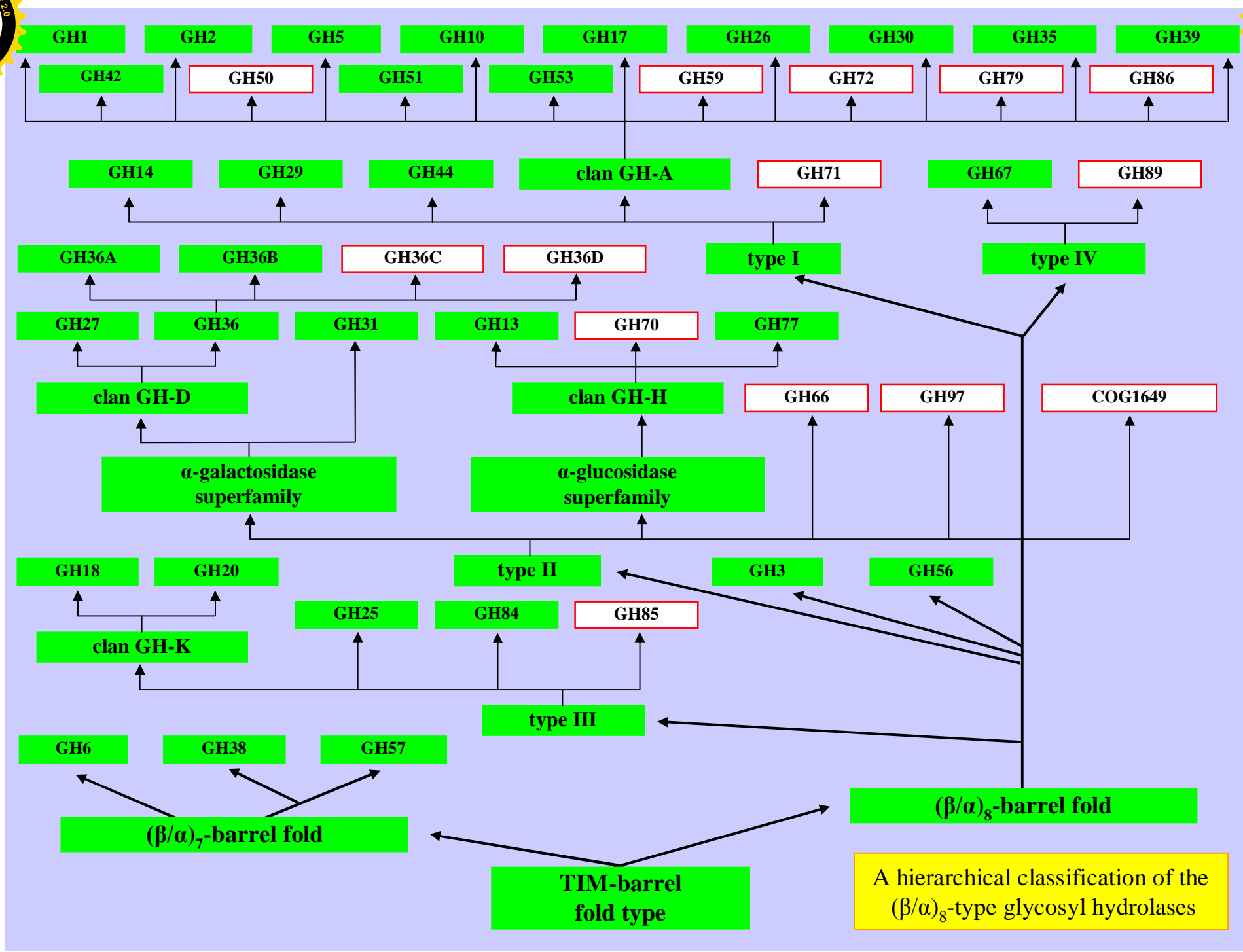
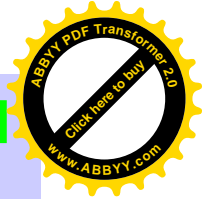
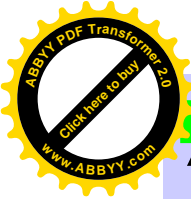


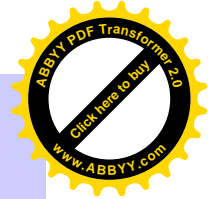
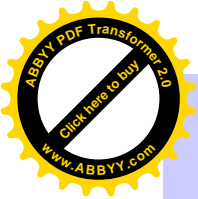


Nagano N, Porter CT, Thornton JM. The $(\beta/\alpha)_8$ glycosidases: sequence and structure analyses suggest distant evolutionary relationships. Protein Eng. 2001, 14(11):845-855.



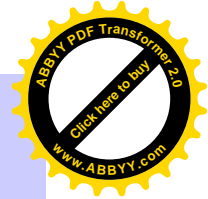
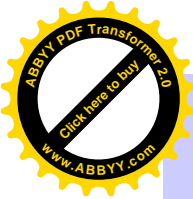
clans: GH-H GH-A ? GH-K





Clans of Glycoside Hydrolases

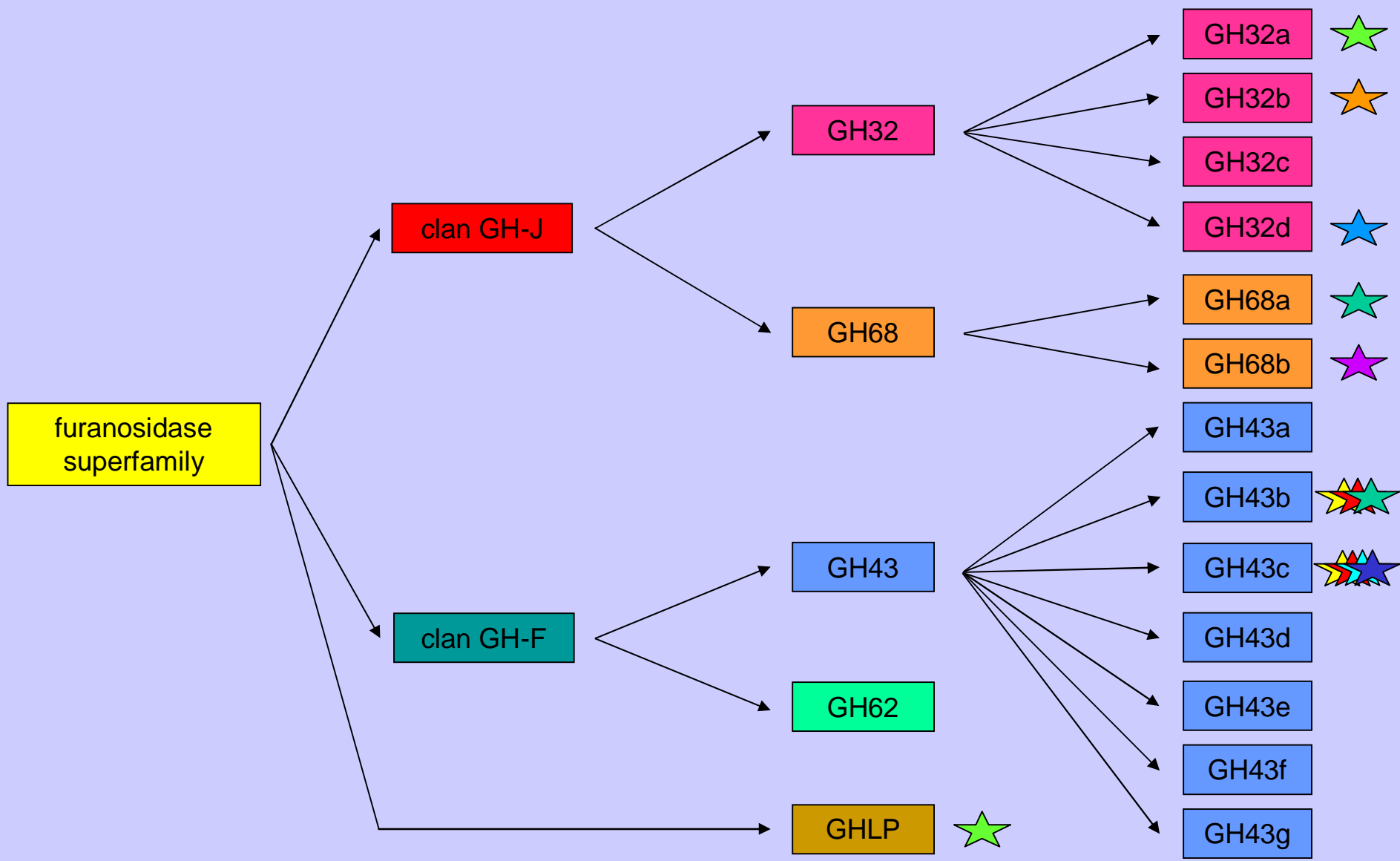
Clan	Families (GH)	Optical Configuration	Tertiary Structure
GH-A	1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 50, 51, 53, 59, 72, 79, 86	retention (equatorial orientation)	$(\beta/\alpha)_8$ -barrel
GH-B	7, 16	retention (equatorial orientation)	β -jelly roll
GH-C	11, 12	retention (equatorial orientation)	β -jelly roll
GH-D	27, 36	retention (axial orientation)	$(\beta/\alpha)_8$ -barrel
GH-E	33, 34, 83	retention (equatorial orientation)	6-fold β -propeller
GH-F	43, 62	inversion (equatorial orientation)	5-fold β -propeller
GH-G	37, 63	inversion (axial orientation)	Not known
GH-H	13, 70, 77	retention (axial orientation)	$(\beta/\alpha)_8$ -barrel
GH-I	24, 46, 80	inversion (equatorial orientation)	$\alpha+\beta$
GH-J	32, 68	retention (β -furanoside)	5-fold β -propeller
GH-K	18, 20	retention (equatorial orientation)	$(\beta/\alpha)_8$ -barrel
GH-L	15, 65	inversion (axial orientation)	$(\alpha/\alpha)_6$
GH-M	8, 48	inversion (equatorial orientation)	$(\alpha/\alpha)_6$
GH-N	28, 49	inversion (axial orientation)	$(\beta)_3$ -solenoid

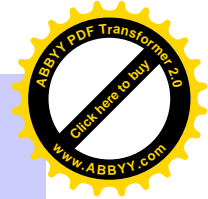
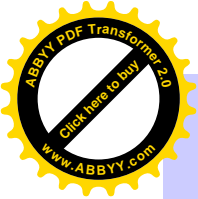


β -fructosidase (furanosidase) superfamily

Family	GH32	GH43	GH62	GH68	GHLP
Clan	GH-J	GH-F	GH-F	GH-J	Not known
COG / KOG	COG1621 / KOG0228	COG3507	None	None	COG2152
Known enzymatic activities	EC 2.4.1.99 EC 2.4.1.100 EC 2.4.1.x EC 3.2.1.7 <u>EC 3.2.1.26</u> EC 3.2.1.65 EC 3.2.1.80	EC 3.2.1.8 EC 3.2.1.37 <u>EC 3.2.1.55</u> EC 3.2.1.99	<u>EC 3.2.1.55</u>	EC 2.4.1.x EC 2.4.1.9 EC 2.4.1.10 <u>EC 3.2.1.26</u>	Not known
Molecular mechanism	Retaining	Inverting	Not known	Retaining	Not known
Origin	Eukaryota: Euglenozoa Fungi Parabasalidea Viridiplantae Eubacteria: Actinobacteria Bacteroidetes Chloroflexi Firmicutes Fusobacteria Proteobacteria Thermotogales Archaea: Euryarchaeota	Eukaryota: Fungi Viridiplantae Eubacteria: Actinobacteria Bacteroidetes Firmicutes Proteobacteria Thermotogales	Eukaryota: Fungi Eubacteria: Actinobacteria Proteobacteria	Eubacteria: Actinobacteria Firmicutes Proteobacteria Archaea: Euryarchaeota	Eukaryota: Fungi Metazoa Viridiplantae Eubacteria: Actinobacteria Aquificales Cyanobacteria Firmicutes Proteobacteria Thermotogales Archaea: Crenarchaeota Euryarchaeota

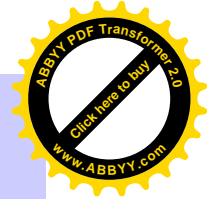
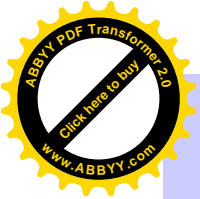
A hierarchical structure of the β -fructosidase (furanosidase) superfamily





Clans of Glycoside Hydrolases

Clan	Families (GH)	Optical Configuration	Tertiary Structure
GH-A	1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 50, 51, 53, 59, 72, 79, 86	retention (equatorial orientation)	$(\beta/\alpha)_8$ -barrel
GH-B	7, 16	retention (equatorial orientation)	β -jelly roll } Strohmeier et al. (2004)
GH-C	11, 12	retention (equatorial orientation)	
GH-D	27, 36	retention (axial orientation)	$(\beta/\alpha)_8$ -barrel
GH-E	33, 34, 83	retention (equatorial orientation)	6-fold β -propeller
GH-F	43, 62	inversion (equatorial orientation)	5-fold β -propeller
GH-G	37, 63	inversion (axial orientation)	Not known
GH-H	13, 70, 77	retention (axial orientation)	$(\beta/\alpha)_8$ -barrel
GH-I	24, 46, 80	inversion (equatorial orientation)	$\alpha+\beta$
GH-J	32, 68	retention (β -furanoside)	5-fold β -propeller
GH-K	18, 20	retention (equatorial orientation)	$(\beta/\alpha)_8$ -barrel
GH-L	15, 65	inversion (axial orientation)	$(\alpha/\alpha)_6$ } ?
GH-M	8, 48	inversion (equatorial orientation)	
GH-N	28, 49	inversion (axial orientation)	$(\beta)_3$ -solenoid



Thank you for the attention!

Please, visit my web-site to find a list of publications:
<http://bioinform.genetika.ru/members/Naumoff/index.htm>