Supplemental Material

Figure S1 Differences between rhomboid subgroups and iRhoms.

Comparative analysis of the subgroups defined in Figure 3A by Two Sample Logo (Vacic et al., 2006) shows amino acids that are enriched and depleted in TMDs 2, 4 and 6 of secretase rhomboids/iRhoms and TMDs 3, 5 and 7 of PARLs (as indicated by position above and below the horizontal line respectively). For each position in the alignment, statistical significance is calculated by a two sample t-test, where the null hypothesis is that the residue is generated according to the same distribution in both positive and negative sample; only residues showing a statistic significant difference (p-value <0.05) are plotted; the size of the symbol is proportional to the observed distribution differences; amino acid residues conserved in the subgroup are indicated in the middle. The position of the artificial splice sites (triangles) and the catalytic motif (GxSx and H) are indicated as in Figure 3A.

Lemberg and Freeman Figure S1

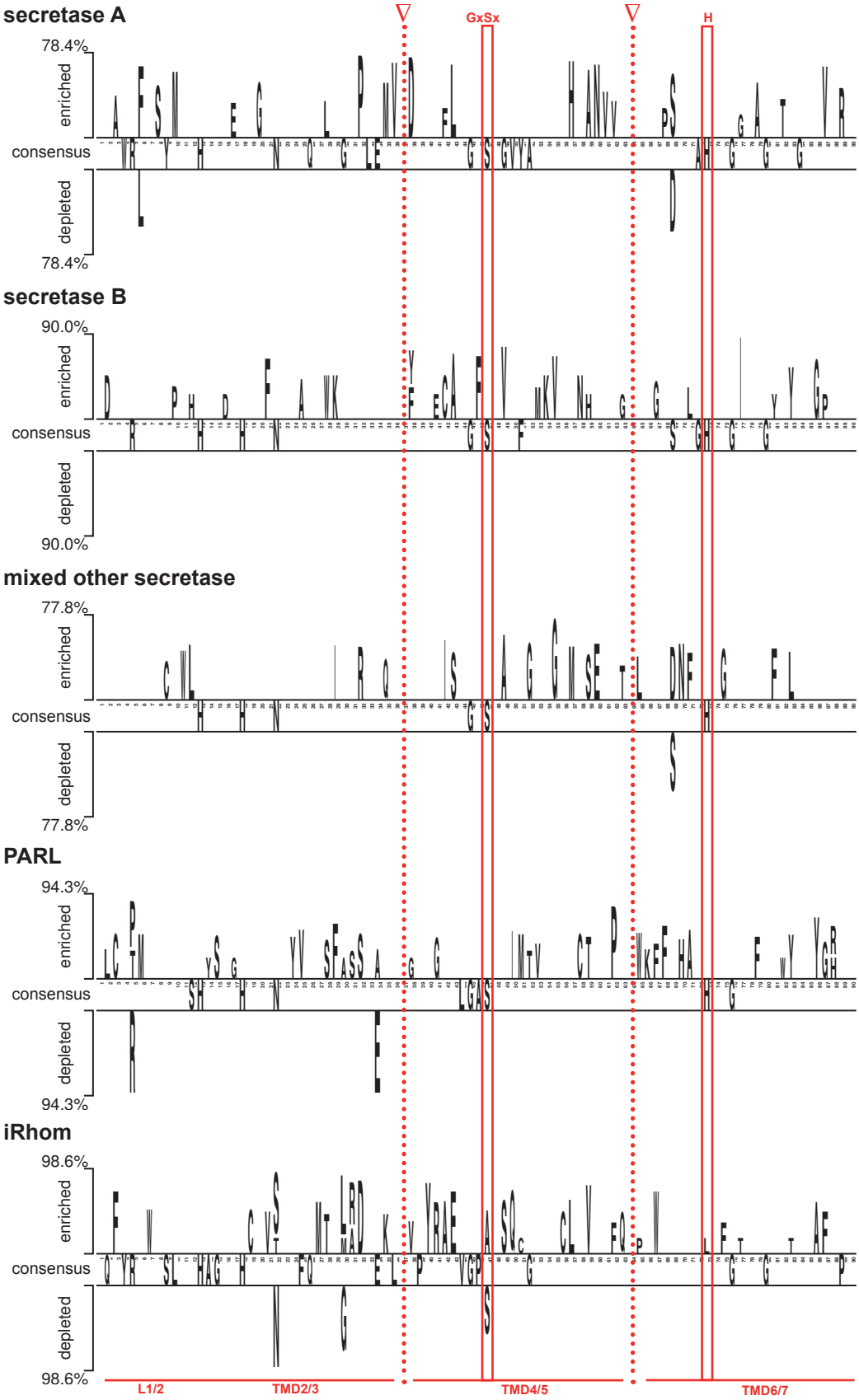

Table S1A Genome-wide analysis of rhomboid homologues in human and mouse reveals that some genes previously annotated as rhomboids are not rhomboid proteases.

Species	Proposed name	Gene ID	Swiss-Prot accession	Synonyms	Basis for proposed name/ consensus mismatch for inactive rhomboid homologues and misannotated genes	automated pfam-annotation/ BLAST-score
Human						
	RHBDL1	9028	O75783	RHBDL, veinlet-like 1, RRP1	published by (Urban et al., 2001); alternative RHBDL, published by (Pascall and Brown, 1998)	pfam01694; 277
	RHBDL2	54933	Q9NX52	veinlet-like 2, RRP2	published (Urban et al., 2001)	pfam01694; 244
	RHBDL3	162494	Q495Y4	ventrhoid, RHBDL4, veinlet-like 3, RRP3, RHBDL3	mouse orthologue published by (Lohi et al., 2004); alternative ventrhoid, published by (Jaszai and Brand, 2002); RHBDL3 preferred for consistency	pfam01694; 164
	RHBDL4	84236	Q8TEB9	RHBDD1	published by (Koonin et al., 2003); named later RHBDD1 by automated annotation; RHBDL4 preferred for consistency	pfam01694; 244
	PARL	55486	Q9H300	PSARL	published by (Pellegrini et al., 2001)	pfam01694; 125
	iRhom1	64285	Q4TT59	RHBDF1	although related to rhomboids, not expected to be active rhomboid protease (see text for details); based on high degree of conservation suggested as new class of iRhoms; alternative RHBDF1 by automated annotation and (Nakagawa et al., 2005); iRhom1 preferred in order to discriminate them from RHBDLs	pfam01694; 274
	iRhom2	79651		RHBDF2, veinlet-like 6	in this study; iRhom2 preferred for consistency; RHBDF2 name was an automatic annotation and this rhomboid has not been the subject of a publication	pfam01694; 259

	distantly related rhomboid homologue	57414	Q6NTF9	RHBDD2, veinlet-like 7, RHBDL7	automated annotation; not predicted to be a rhomboid protease or an iRhom; potentially distant evolutionary relation to rhomboids, however, low sequence similarity (less then 13% identity to sequences shown in Figure 1) *consensus mismatch:* no TMD2-signature; GFTP instead of GxSx in putative TMD4; N instead of H in putative TMD6	pfam01694; 135
	no significant relation to rhomboids and iRhoms	25807	Q9Y3P4	RHBDD3	automated annotation; not predicted to be a rhomboid protease or an iRhom *consensus mismatch:* no TMD2 signature; GLSS in putative TMD4; no H in putative TMD6	no value
Mouse						
	RHBDL1	214951	Q8VC82	veinlet-like 1	published by (Lohi et al., 2004)	pfam01694; 257
	RHBDL2	654339		veinlet-like 2	published by (Urban and Freeman, 2003)	pfam01694; 233
	RHBDL3	246104	P58873	veinlet-like 3	published by (Lohi et al., 2004); alternative ventrhoid, published by (Jaszai and Brand, 2002). RHBDL3 preferred for consistency	pfam01694; 195
	RHBDL4	76867	Q8BHC7	Rhbdd1	in this study; alternative Rhbdd1 by automated annotation; RHBDL4 preferred for consistency	pfam01694; 244
	PARL	381038	Q5XJY4	PSARL	Published by (Cipolat et al., 2006); orthologue to human PARL	pfam01694; 219
	iRhom1	13650	Q6PIX5	RHBDF1	in this study; orthologue to human iRhom1 (e.g. indicated by characteristic LAG-motif in the variable region of the IRHD; see Figure 5)	pfam01694; 275

iRhom2	217344	Q80WQ6	RHBDF2, rhomboid-like protein 6	in this study; orthologue to human iRhom2 (e.g. indicated by characteristic DKSD-motif in the variable region of the IRHD; see Figure 5)	pfam01694; 294
distantly related rhomboid homologue	215160		Rhbdd2, RHBDL7, Usmg1	automated annotation; not predicted to be a rhomboid protease or an iRhom; potentially distant evolutionary relation to rhomboids, however, low sequence similarity only (less then 13% identity to sequences shown in Figure 1) *consensus mismatch:* no TMD2-signature; GFTP instead of GxSx in putative TMD4; N instead of H in putative TMD6	pfam01694; 161
no significant relation to rhomboids and iRhoms	279766		Rhbdd3	automated annotation; not predicted to be a rhomboid protease or an iRhom *consensus mismatch:* no TMD2 signature; GLSG in putative TMD4; no H in putative TMD6	no value

Table S1B Genome-wide analysis of rhomboid homologues in *Arabidopsis thaliana* reveals that previously annotated genes are no rhomboid proteases.

Species	Proposed name	Gene ID	locus	Synonyms	Basis for proposed name/ consensus mismatch for inactive rhomboid homologues and misannotated genes	automated pfam-annotation/ BLAST-score
Arabidopsis						
	RBL1	817453	At2g29050		published by (Kanaoka et al., 2005)	pfam01694; 179 COG0705; 170
	RBL2	842616	At1g63120		published by (Kanaoka et al., 2005)	pfam01694; 189
	RBL3	830616	At5g07250		published by (Koonin et al., 2003); name suggested by (Kanaoka et al., 2005)	pfam01694; 162
	RBL4	824545	At3g53780		published by (Kanaoka et al., 2005)	pfam01694; 333
	RBL5	841690	At1g52580		published by (Kanaoka et al., 2005)	pfam01694; 219
	RBL6	837831	At1g12750		published by (Kanaoka et al., 2005)	pfam01694; 292
	RBL7	828406	At4g23070		published by (Kanaoka et al., 2005)	pfam01694; 279
	RBL8	839113	At1g25290	RBL10	published by (Koonin et al., 2003); alternative RBL10, published by (Garcia-Lorenzo et al., 2006); RBL8 preferred for consistency	pfam01694; 294
	RBL9	832642	At5g25752	RBL11	published by (Koonin et al., 2003); alternative RBL11, published by (Garcia-Lorenzo et al., 2006); RBL9 preferred for consistency	pfam01694; 187
	RBL10	821028	At3g17611	RBL14	published by (Koonin et al., 2003); alternative RBL14, published by (Garcia-Lorenzo et al., 2006); RBL10 preferred for consistency	pfam01694; 124

	RBL11	825015	At3g58460	RBL15	published by (Koonin et al., 2003); alternative RBL15, published by (Garcia-Lorenzo et al., 2006); RBL11 preferred for consistency	pfam01694; 196
	RBL12	825121	At3g59520	RBL13	published by (Koonin et al., 2003); alternative RBL13, published by (Garcia-Lorenzo et al., 2006), RBL12 preferred for consistency	pfam01694; 163
	PARL	838441	At1g18600	RBL12	published by (Koonin et al., 2003); alternative RBL12, published by (Garcia-Lorenzo et al., 2006), PARL preferred for consistency	pfam01694; 139
	inactive rhomboid homologue unrelated to iRhoms	844122	At1g77860	KOMPEITO, RBL8	suggested by (Koonin et al., 2003) and by (Garcia-Lorenzo et al., 2006); although related to rhomboid not predicted to be a rhomboid protease or an iRhom *consensus mismatch:* SSGA instead GxSx in TMD4; N instead of H in TMD6	pfam01694; 266
	inactive rhomboid homologue unrelated to iRhoms	833839	At5g38510	RBL9	suggested by (Tripathi and Sowdhamini, 2006) and by (Garcia-Lorenzo et al., 2006); although related to rhomboid not predicted to be a rhomboid protease or an iRhom *consensus mismatch:* GGTG instead of GxSx in TMD4; N instead of H in TMD6	pfam01694; 297
	inactive rhomboid homologue unrelated to iRhoms	843753	At1g74130		suggested by (Tripathi and Sowdhamini, 2006) and (Karakasis et al. 2007); although highly related to At1g18600 (PARL) not predicted to be a rhomboid protease or an iRhom *consensus mismatch:* GANG instead of catalytic GxSx (GGSV in TMD3 is in topological incorrect position); Q instead of catalytic H	pfam01694; 124
	inactive rhomboid homologue unrelated to iRhoms	843754	At1g74140		suggested by (Tripathi and Sowdhamini, 2006); although related to At1g18600 (PARL) not predicted to be a rhomboid protease or an iRhom *consensus mismatch:* GADG instead of catalytic GxSx (GGSV in predicted TMD3 is in topological incorrect position); Q instead of catalytic H	pfam01694; 116

distantly related rhomboid homologue	819986	At3g07950	suggested by (Tripathi and Sowdhamini, 2006); not predicted to be a rhomboid protease or an iRhom; potentially distant evolutionary relation to rhomboids, however, low sequence similarity only (less then 15% identity to sequences shown in Figure 1) *consensus mismatch:* no TMD2-signature; GLLVG instead of GxSx in TMD5; no H in TMD7	pfam01694; 290
distantly related rhomboid homologue	824841	At3g56740	suggested by (Tripathi and Sowdhamini, 2006); not predicted to be a rhomboid protease or an iRhom; potentially distant evolutionary relation to rhomboids, however, low sequence similarity only (less then 17% identity to sequences shown in Figure 1) *consensus mismatch:* no TMD2-signature; GPYG instead of GxSx in TMD4; no H in TMD6	no value
no significant relation to rhomboids and iRhoms	818715	At2g41160	suggested by (Tripathi and Sowdhamini, 2006); not predicted to be a rhomboid protease or an iRhom *consensus mismatch:* no TMD2-signature; GPYA instead of GxSx in TMD4; no H in TMD6	no value
no significant relation to rhomboids and iRhoms	818492	At2g39060	suggested (Tripathi and Sowdhamini, 2006); not predicted to be a rhomboid protease or an iRhom *consensus mismatch:* no TMD2-signature; GWVC instead of GxSx in TMD5; no H in TMD7	no value
no significant relation to rhomboids and iRhoms	832640	At5g25640	automated annotation; not predicted to be a rhomboid protease or an iRhom *consensus mismatch:* no TMD2-signature; GASG in TMD2 (of 3 predicted TMDs)	pfam01694; 80

Table S2 Topology prediction of PARLs across evolution.

Predicted TMDs ignored in the multiple sequence alignment shown in Figure 2A are highlighted in italic font. In summary, combining the various results increases the quality of the topology model (Nilsson et al., 2000), indicating that the seven TMD-topology (1+6, see text for details) is conserved among PARLs. Even for the more distant plant orthologues a 1+6 TMD structure is likely. See Figure 3A and Table S1 for protein accession numbers.

Species		**TMD1**	**TMD2**	**TMD3**	**TMD4**	**TMD5**	**TMD6**	**TMD7**
consensus						GxSx		H
Hs PARL								
	TMHMM	102-121	168-187	208-230	245-262	-	296-318	*333-354*
	HMMTOP	102-119	170-188	219-236	245-263	272-289	298-317	*336-353*
	PSORT	-	169-185	-	-	-	-	-
	TMpred	101-119	168-188	*222-242*	245-263	273-289	296-318	*329-348*
Mm PARL								
	TMHMM	100-119	166-185	*206-228*	243-260	-	294-316	*331-352*
	HMMTOP	100-117	168-186	217-234	243-261	270-287	296-315	*334-351*
	PSORT	-	-	-	-	-	294-310	-
	TMpred	99-117	166-186	220-240	243-261	271-287	294-316	327-346
Dm Rhomboid-7								
	TMHMM	73-95	148-167	180-202	-	244-266	275-297	312-333
	HMMTOP	77-95	148-167	198-217	224-241	250-268	275-294	*315-333*
	PSORT	-	-	-	225-241	-	-	-
	TMpred	77-95	147-167	180-199	*(201-220)* 226-242	249-265	275-297	312-332
Ce ROM5								
	TMHMM	-	-	-	173-190	*197-219*	*234-253*	260-282
	HMMTOP	71-88	-	*125-142-*	173-190	201-219	*236-253*	264-281
	PSORT	-	-	-	-	-	*239-255*	-
	TMpred	68 -86	-	132-161	172-190	201-217	*238-258*	262-281
Tg ROM6								
	TMHMM	-	-	-	-	-	-	-
	HMMTOP	275-294	*307-325*	356-375	392-410	441-460	-	-
	PSORT	281-297	-	-	-	-	-	-
	TMpred	*(222-246)* 279-297	*307-326*	369-385	391-410	*440-468*	-	488-506

Sc Pcp1/Rbd1								
	TMHMM	107-129	144-163	*204-226*	-	246-268	275-297	310-327
	HMMTOP	*102-124*	145-162	*199-219*	-	246-269	276-295	314-331
	PSORT	-	-	-	-		275-291	-
	TMpred	106-129	145-163	199-219	212-241	246-270	275-303	308-326
At PARL								
	TMHMM	-	*135-153*	*178-200*	*213-235*	-	*275-297*	-
	HMMTOP	-	*135-154*	*187-205*	*218-237*	-	*278-297*	*310-327*
	PSORT	-	*134-150*	-	-	*256-272*	*280-296*	-
	TMpred	-	*135-153*	*187-205*	*216-237*	*253-271*	*278-297*	*310-327*
Os Os01g55740								
	TMHMM	-	*111-130*	-	*185-207*	*228-247*	*251-270*	-
	HMMTOP	-	*111-130*	-	*188-207*	*228-246*	*251-270*	*283-302*
	PSORT	*60-76*	*111-127*	-	*190-206*	*230-246*	*254-270*	-
	TMpred	*61-81*	*111-127*	*167-191*	*193-213*	*229-247*	*255-275*	*283-301*

Table S3 Predicted structural motifs in tails of eukaryotic rhomboids.
See Methods for details, Figure 3A and Table S1 for protein accession numbers.
Certain rhomboids have predicted structural motif in their N- and C-terminal tails (Koonin et al., 2003). The function of these cytosolic putative domains is not clear and we therefore have not used them as a basis for classification. The certainty of prediction is low in some cases (indicated by high E-values).

Domain	**Pfam accession**	**Position**	**Protein**	**Species distribution**	**E-value**
EF-hand	cd00051	N-terminal	RHBDL1	vertebrates	1.3e-3
			RHBDL3	vertebrates	2e-5
			Rhomboid-4	*Drosophila*	1e-7
Zn-finger domains	pfam00641	C-terminal	At RBL10	Plants	1e-3
	smart00547		Os01g18100		6e-4
Ubiquitin associated domain	cd00194	C-terminal	At RBL11	Plants	low similarity
			Os03g44830		8e-5

Table S4 Sequence relationship of PARL-type rhomboids across evolution. Pairwise alignment using the EMBOSS algorithm reveals the degree of sequence identity and similarity (compared to human PARL). MEME-MAST analysis illustrates a statistically valid homology between the 11 tested PARL orthologues that is based on characteristic sequence motifs defined by MEME (10 motifs with best possible match, not shown); the E-value of a sequence is a similarity score, expressing the expected number of sequences in a random database of the same size that would match the characteristic motifs at the observed level (Bailey and Gribskov, 1998). See Figure 3 and Table S1 for accession numbers.

Species	Length	% identity/similarity (EMBOSS)	E-value (MAST)
human	379	100	0
mouse	377	90.5/94.5	0
D. rerio	383	52.1/65.9	4.3e-214
D. melanogaster	351	37.2/55.1	1e-214
D. pseudoobscura	354	39.9/56	2.7e-211
C. elegans	300	29.6/42.3	2.4e-93
S. cerevisiae	346	21.3/33.7	2.4e-56
A. thaliana	336	21.2/31.1	1.7e-34
O. sativa	307	22.1/32.1	3.6e-29
P. falciparum	569	13.7/26.0	1.4e-12
T. gondii	641	12.2/19.7	0.081

Table S5 Sequence relationship of secretase B-type rhomboids to RHBDL4 orthologues. Pairwise alignment using the EMBOSS algorithm reveals the degree of sequence identity and similarity (compared to human RHBDL4). MEME-MAST analysis illustrates a statistically valid homology between the 10 tested B-type rhomboids that is based on characteristic sequence motifs defined by MEME (see Table S4 for details of the analysis). The RHBDL4 subgroup is characterised by the operational orthology definition of bi-directional best hit in a Smith-Waterman comparison (result not shown) (Snel et al., 1999). See Figure 3 and Table S1 for accession numbers.

Species/Name	**Length**	**% identity/similarity (EMBOSS)**	**E-value (MAST)**
RHBDL4s			
human RHBDL4	315	100	4.9e-228
mouse RHBDL4	315	81.0/88.9	1.7e-236
D. rerio RHBDL4	335	56.2/70.5	7.9e-180
A. thaliana RBL10	334	31.2/44.3	2.9e-108
O. sativa Os01g18100	350	29.6/41.1	2.9e-110
others			
A. thaliana RBL11	402	17.8/28.7	3.4e-115
O. sativa Os03g44830	399	19.5/30.7	3.9e-113
A. thaliana RBL12	240	24.0/37.1	3.3e-56
O. sativa Os01g67040	273	18.4/28.0	3e-55
S. cerevisiae Rbd2	262	18.8/30.3	3.4e-27

Table S6 Sequence relationship of iRhoms across evolution. Pairwise alignment using the EMBOSS algorithm reveals the degree of sequence identity and similarity (compared to human iRhom1). MEME-MAST analysis illustrates a statistically valid homology between the 21 tested iRhoms that is based on characteristic sequence motifs defined by MEME (see Table S4 for details of the analysis). See Figure 5 and Table S1 for accession numbers. Note that the genes of *A. gambiae* iRhom and *C. intestalis* iRhom are only partially predicted and their sequences lack the more variable N-terminal domain (and TMD1 for *C. intestalis*).

Species/Name	**Length**	**% identity/similarity (EMBOSS)**	**E-value (MAST)**
human iRhom1	855	100	0
human iRhom2	856	56.0/68.0	0
P. troglodytes iRhom1	862	99.1/99.1	0
P. troglodytes iRhom2	856	55.4/67.5	0
C. familiaris iRhom1	856	95.1/97.2	0
C. familiaris iRhom2	855	55.8/67.6	0
B. taurus iRhom1	856	94.6/97.5	0
B. taurus iRhom2	830	56.4/69.1	0
mouse iRhom1	856	95.3/97.8	0
mouse iRhom2	827	57.0/69.6	0
R. norvegicus iRhom1	856	95.2/97.5	0
R. norvegicus iRhom2	825	57.1/69.7	0
G. gallus iRhom1	769	77.9/84.7	0
G. gallus iRhom2	691	46.5/58.1	1.3e-293
D. rerio iRhom	857	80.2/89.9	0
T. rubripes iRhom	894	78.1/86.2	0
C. intestinalis iRhom	*(430, partial seq.)*	*26.3/33.2*	*3.2e-178*
D. melanogaster iRhom (Rhomboid-5)	700	22.9/31.9	1.8e-161
A. gambiae iRhom	*(422 partial seq.)*	*23.3/32.8*	*1.2e-158*
C. elegans iRhom1 (ROM3)	861	27.4/40.6	1e-166
C. elegans iRhom2 (ROM4)	727	24.4/35.6	1.4e-147

References

Bailey, T.L. and M. Gribskov. 1998. Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* **14**: 48-54.

Cipolat, S., T. Rudka, D. Hartmann, V. Costa, L. Serneels, K. Craessaerts, K. Metzger, C. Frezza, W. Annaert, L. D'Adamio, C. Derks, T. Dejaegere, L. Pellegrini, R. D'Hooge, L. Scorrano and B. De Strooper. 2006. Mitochondrial rhomboid PARL regulates cytochrome c release during apoptosis via OPA1-dependent cristae remodeling. *Cell* **126**: 163-175.

Garcia-Lorenzo, M., A. Sjodin, S. Jansson and C. Funk. 2006. Protease gene families in Populus and Arabidopsis. *BMC Plant Biol* **6**: 30.

Jaszai, J. and M. Brand. 2002. Cloning and expression of Ventrhoid, a novel vertebrate homologue of the Drosophila EGF pathway gene rhomboid. *Mech Dev* **113**: 73-77.

Kanaoka, M.M., S. Urban, M. Freeman and K. Okada. 2005. An Arabidopsis Rhomboid homolog is an intramembrane protease in plants. *FEBS Lett* **579**: 5723-5728.

Koonin, E.V., K.S. Makarova, I.B. Rogozin, L. Davidovic, M.C. Letellier and L. Pellegrini. 2003. The rhomboids: a nearly ubiquitous family of intramembrane serine proteases that probably evolved by multiple ancient horizontal gene transfers. *Genome Biol* **4**: R19.

Lohi, O., S. Urban and M. Freeman. 2004. Diverse substrate recognition mechanisms for rhomboids; thrombomodulin is cleaved by Mammalian rhomboids. *Curr Biol* **14**: 236-241.

Nakagawa, T., A. Guichard, C.P. Castro, Y. Xiao, M. Rizen, H.Z. Zhang, D. Hu, A. Bang, J. Helms, E. Bier and R. Derynck. 2005. Characterization of a human rhomboid homolog, p100hRho/RHBDF1, which interacts with TGF-alpha family ligands. *Dev Dyn* **233**: 1315-1331.

Nilsson, J., B. Persson and G. von Heijne. 2000. Consensus predictions of membrane protein topology. *FEBS Lett* **486**: 267-269.

Pascall, J.C. and K.D. Brown. 1998. Characterization of a mammalian cDNA encoding a protein with high sequence similarity to the Drosophila regulatory protein Rhomboid. *FEBS Letters* **429**: 337-340.

Pellegrini, L., B.J. Passer, M. Canelles, I. Lefterov, J.K. Ganjei, B.J. Fowlkes, E.V. Koonin and L. D'Adamio. 2001. PAMP and PARL, two novel putative metalloproteases interacting with the COOH-terminus of Presenilin-1 and -2. *J Alzheimers Dis* **3**: 181-190.

Snel, B., P. Bork and M.A. Huynen. 1999. Genome phylogeny based on gene content. *Nat Genet* **21**: 108-110.

Tripathi, L.P. and R. Sowdhamini. 2006. Cross genome comparisons of serine proteases in Arabidopsis and rice. *BMC Genomics* **7**: 200.

Urban, S. and M. Freeman. 2003. Substrate specificity of rhomboid intramembrane proteases is governed by helix-breaking residues in the substrate transmembrane domain. *Mol Cell* **11**: 1425-1434.

Urban, S., J.R. Lee and M. Freeman. 2001. Drosophila rhomboid-1 defines a family of putative intramembrane serine proteases. *Cell* **107**: 173-182.

Vacic, V., L.M. Iakoucheva and P. Radivojac. 2006. Two Sample Logo: a graphical representation of the differences between two sets of sequence alignments. *Bioinformatics* **22**: 1536-1537.