# Supplementary Materials: The Distribution of Lectins across the Phylum Nematoda: A Genome-Wide Search

## Lander Bauters, Diana Naalden and Godelieve Gheysen



**Figure S1.** Alignment of partial calreticulin/calnexin sequences. Amino acids are represented by one letter codes in different colors. Residues needed for carbohydrate binding are indicated in red boxes. Sequences containing all six necessary residues are indicated with an asterisk.

#### Int. J. Mol. Sci. 2017, 18, 91; doi:10.3390/ijms18010091

EcorL	IWDMTTG-TVASFETRESESIEQPYTRPLPADGLVEEMGP	TKSKPAQGYGYLGIENNSKODNSYOTLGVEEDT	SNPWDPPQVPHIGID VNSIRS-IK TOPFQLDNG
GPLIN_000532200	* AWN KRAFTESDH FELQFAFKVTGQ GRVGGDGLAIWYTA	QQGNLGPVFGA NDFWTGMAIMLDS	DNDGARNNPAISLMINDGFVLG
BUX.s01198.147	* AWN KRAYTEAEN FEVD VAFKITGQ RIGADGLAVWYTA	SQGTLGPVF <mark>G</mark> A NDHWNGLAIMFDS	D N D G R K D N P Y V S V M V N D G T K A Y N H H N D G S D Q I L G -
K07A1.8	* AWN KRAFLESEN FQVD I ALKIGGQ GRVGADGLGIWYTS	5 Q L G A L G P V F <mark>G</mark> G N D F W T G M G L F L D S I	DNDG QKNNPQVSLMLNDGTRSYDHHTDGSQQILS -
Bm4849b	* AWN KRVMAESSHFEVQVAFRVTGQ GRIGADGIAIWYTA	QQPTLGPVFGANDYWIGMGLFLDS	DNDA QKNNPFVALMINDGTRQYDHQTDGSQQMLS -
Mh10g200708_Contig1358_6198.	. * AWN K R Q F L A S D H F E L E V V F K V T G Q G R I G A D G L A I WY T A	QQGT I GP V F G A N D F WT G M G I M L D S	DINDGAKNNPTISLMINDGIF
GS_11102	* AWNKRPMIESENFEIEIAFKITGQ GRIGADGLAVWYTA	QQGTLGPVFGSNDFWTGMGLFFDS	DINDG QKNNPFVALMINDGTRQYDHQTDGSQQMLS -
GPLIN_000532100	* AWN KRAFTESDHFELQFAFKVTGQ GRVGGDGLAIWYTA	QQGNLGPVFGANDFWTGMAIMLDS	DINDG ARNNPAISLMINDGTRAFDHPTDGSHQVLS -
GS_07261	MRIASPI - PFRSWEVEITFKIYGS - DAGQSGEGMAFWYVE	QSTRRGRAFGFPDVFRGLGVFIDTS	EDDVTDINHYHPFISALVNNGTIQYAHDAFGTYSQLGG
BUX.s00351.105	* IWSQIPV - QSRDWELMVTFRIHGD - TGKLFGDGIGIWYTR	R D R N I P G P V F G S T N G F S G L G V F L D T	H N D Y S S Y A H T F P Y I Y A M V S D G T V L F N G D K D G G D D R L G G
GPLIN_000771000	LWNKVPN-SVRDWEMVVSFKVHGS-TGALFGG	P G N F F R G L G I F L D T	SNHNGPHQHAHPYISAMVSNGTLQYDHDADGTHTQLGG
gnl MincDB prot:Minc18157	* I WN K V P N - H A R D W E L V V D F R V Y G S - T G T L F G D G F A I W Y V Q	DPNIVGPVFGSKDYFRGLGIFLDT	SNHNGPHSHSHPFISAMVSNGSLHYDHDKDGTHTQLGG
Mh10g200708_Contig1149_2451	7 * I WN K V P N - H A R D W E L V V D F R V H G S - T G S L F G D G F A I W Y V Q	EPNIPGPVFGSKDYFRGLGIFLDT	SNHNGPHAHSHPFISSMVSNGSLHYDHDKDGTHTQLGG
T04G9.3	* LWN T Q P V - WS R D W E L Q V S F K V T G S - T G D L F G D G M A I W Y T S	EPNILGPVFGGKDYFRGLAVFLDT	SNHNGPHQHGHPFISAMVSDGSLHYDHDKDGTHTQLGG
GS_08081	* LWN TMP V - WS R DWELL VN FK VHGT - TGDL FGDG FAIWY VQ	ERSQAGNVFGSKDFFRGLAIFLDT	SNHNGPHSHGHPYISAMVNNGSLHYDHDKDGTHTQLGG
Bm5558	* IWN N V P V - N V R D W E L H V N F A I H G S - T G D L F G D G A A I W Y V Q	DPAQTGPVFGSKDYFRGLGIFLDT	SNHNGPHEHGHPYISAMINNGSLHYDHDMDGTHTQLGG

**Figure S2.** Alignment of partial legume lectin-like sequences. Amino acids are represented by one letter codes in different colors. EcorL is a legume lectin originating from *Erythrina corallodenron,* used in this alignment to compare carbohydrate binding sites. The residues necessary for carbohydrate interaction are shown in red boxes. Nematode lectin-like sequences containing at least four out of five key residues are indicated with an asterisk.



**Figure S3.** Alignment of possible Ricin-B lectin-like domains. Amino acids are represented by one letter codes in different colors. The key amino acid residues (D-Q-W) involved in carbohydrate binding, which are repeated three times, are boxed in red. Sequences that have at least one complete D-Q-W triad are indicated with an asterisk.

gnl\MincDB prot:Minc15214	HNVKNGESLMQISLLYSVPVSEIKRVNNIVADQEIHAFPVIKVP
Mh10g200708_Contig849_13692.	HNVKNGETLMQISLLYSVPVSEIKRVNNIVADQEIHAFPVIKVP
F43G9.2b	- KVKNGDTLTTKLAIKYQVNVAEIKRVNNMVSEQDFMAL
F43G9.2a	- K V K N G D T L N K L A I K Y Q V N V A E I K R V N N M V S E Q D F MAL
Bm4065b	- KLRPGDNLNKIAVQYSVSLSDLKRVNNIVSEQDIYAMPFVKIP
Bm4065a	- KLRPGDNLNKIAVQYSVSLSDLKRVNNIVSEQDIYAMPFVKIP
Mh10g200708_Contig1499_16049	VEVQNTDILEGIAAAHDCTVGFLVKLNKLTSRM-VFPGQKLLVP
Bm4263g	YTVAQSDSTERIAAVHDCTVGELMKMNRLATRM-VFPGQKLLVP
GS_02371	YMVKDHD_SILERIAAFHDCTVGELMKMNRMGSRM-VFPGQKLMVP
F52E1.13a	YTITETDTLERVAASHDCTVGELMKLNKMASRM-VFPGQKILVP
F52E1.13c	YTITETDTLERVAASHDCTVGELMKLNKMASRM-VFPGQKILVP
F52E1.13d	YT I T ET D T L ER VAASH D CT V G EL MKL N KMASRM - V F P G
B0041.3	YQVQTDDTLERIALKHNCSVSSLVRANKLWSPSALFMKQFIRIP
BUX.s00713.252	HEVKPTDTLQNLQIKYNSNMFEIKRINKLWSNDALYAR
gnl MincDB prot:Minc17986	HKVQPNDTLQNLELRYNSSMFEIKRINRLWSNQSLYCKTEILIP
Mh10g200708_Contig1798_19313	HKVQPNDTLQNLELRYNSSMFEIKRINRLWSNQSLYCK
GS_14918	HQVEPTDTLQGLVLRYNTSMSEIKRLNRLWSNESLYLKPYINIP
Bm1699b	HQIKPNDTLQGLVLKYNTSMSEIKRLNRLWSNESLYLKEYIEIP
F07G11.9_11	AT I T P G N T C F N I S VAY G I N L T D L Q K T Y D C KAL E V G D T I C V S
T01C4.1_3	ATVTPGSSCYTISASYGLNLAELQTTYNCDALQVDDTICVS
Mh10g200708_Contig1786_1647.	HKIKAGDSCEGTALKYGISIEQLYKINHWNAEKCKKLAIF
gnl MincDB prot:Minc14493	GISIEQLYTINHWNAEKCKRLATGDFICVA
F07G11.9_12	IEVIKGDTCMFLENAFKTNQTEMERANEGVKCDNLPIGRMMCVW
T01C4.1_4	IEILNGDTCGFLENAFQTNNTEMEIANEGVKCDNLPIGRMMCVW
F07G11.9_1	TEIKSGDSCMNIASNAKISVERLQQLNKGMKCDKLPLGDKLCLA
T01C4.1_1	HTIKSGDTCMKIASEASISVQELEGLNSKKSCANLAVGLSEQEFTEMN EELDCNRLAVGNEICVS
F07G11.9_5	YRFKKGDTCYKIWTSHKMSEKQFRALNRGIDCDRLVPGKELCVG
F07G11.9_3	HKIQKGDTCFKIWTTNKISEKQFRNLNKGLDCDKLEIGKEVCIS
F07G11.9_4	LKIKEGDTCYNIWTSQKISEQEFMELNKGLDCDKLEIGKEVCVT
T01C4.1_2	IHVKEGDTCYTIWTSQHLTEKQFMDMNEELNCGMLEIGNEVCVD
F07G11.9_2	- KLKAEDTCFKIWSSQKLSERQFLGMNEGMDCDKLKVGKEVCVA
F07G11.9_9	TEVKEGDTCFKIWSAHKITEQQFMEMNRGLDCNRLEVGKEVCIV
F07G11.9_10	IKVKEGDTCFKIWSAQKMTEQQFMEMNRGLDCNKLMVGKEVCVS
F07G11.9_8	VQINPGDTCFKIWSAQ
F07G11.9_6	ITVKPGDTCFSIWTSQKMTQQQFMDINPELDCDKLEIGKEVCVT
F07G11.9_7	VK I N P G D C EN I WT S Q
	* * *

**Figure S4.** Alignment of possible LysM lectins. Amino acids are represented by one letter codes in different colors. Conserved cysteine residues are marked with an asterisk under the alignment. The key residue involved in carbohydrate binding in an eukaryote is boxed in red [1]. The conserved key residues characterized in bacterial sequences are boxed in blue [2].



**Figure S5.** Alignment of potential M-type lectins. Amino acids are represented by one letter codes in different colors. M-type lectin-like sequences are compared with a human alpha-mannosidase (Q9UKM7.2). Conserved cysteine residues in M-type lectin-like sequences are denoted with an asterisk under the alignment. Key residues necessary for mannosidase activity are boxed in red. These residues are absent in M-type lectins.

	the state of the s		
Q9BZP6.1	IKFLRQYEFDGL	DIDNE	YPGS-RGSPPQDKHLFTVLVQEMREAFEQEAK
Bm10739a	I A F L R K N N F D G F	DLDNE	YPVG-VAEEHAKLVEAMKTAFVEEAK
TOICAI	ASELENHKIDGY	DIDAE	PENKOG DKONHVTLIREIRERENGMAS
Toroact			PERCONTENT OR ANTIFIC TREERENTION AS
113H5.3	VKVIKEYGFDGV	DIDNE	TPVIGGAVEG TPADRKNYVNLMRELKNELKDLES
GPLIN_000707700	NG LDG A	DIDNE	FPTANDKANYVSFLSELKKALPR
Bm9336	I A FLRKNN FDG F	DLDVE	YPVG-VAEEHAKLVE
Em 917/3	LAFLEWNNEDGE	DIDUE	PVC - V AFEHAKI VEAMATAEVEEAK
Destantes	LAT LARANT DO		
5/1951a	I AF LKKNN FDG F	DIDVE	YPVG-VAEEHAKLVEAMKIAFVEEAK
Y22D7AL.14	TQVLQKYRLNGV	DIDNE	FPVWSR DAQKSDKAHFATFLRILKSHLKPA
BUX 501038 118	VAFAKKNG FDGV	DIDAF	FPTA . F OK SEETOLIKAVK SEAG
DUN ANAAA AT	WARAWONG FROM		
BUX.501661.2/	VAFAKQTGFDGV	DIDNE	PPIA-DQKDLFIQLVQAVKIAGG
BUX.600609.89	VD FAKHHG FDG V	DIDNE	FPSG-AQKHGFTELIKAVKAAAG
BUX s00351.397	RDYLI SNQLDGF	DIDVE	FPAFSP DAKPSDRDGFTFLLKALREKFDQSG -
BUX c01002 2	VE FARRIG FOCH	DIDDE	EDTA-COKDI ETOLIKAVKOAAC
DUX.SVIVEZZ	VET ARRIED DO V		TELEVISION CONCEPTION AND AND AND AND AND AND AND AND AND AN
GS_04351	TOYIVKKSVDG	DIDNE	PPVVVSA DARPTOKKGPTVLLEELKKSPDKAR -
Bm8330f	TRYLVANELSGF	DIDNE	FPVWSR DAQPTDRKGLSVLIKELRESFDQAK -
GPI IN 000582100	VSFLFSNGLDG	DIDME	FRTVN DKANYVAFILFINKALPS
anil Mine O Player, Mine 07/89	KKNC FOCI		ENENLILOFEDIALEN
gnifmineDB[proc:mine07463	KKNGPDGI		TPD S-KERFRELLUEFRLASSREK-
gni[MincDB prot:Minc15121	I A F I KKHG FDG I	DLDIE	YPDS-KENFNLLLQEFHLASCNEK-
GS 08584	I A F V R K Y N F D G I	DIDVE	YPTG - P QDKENFSAFLKNERVV SDGHRR
CC 15811	I A EVEK YN EDG I	DIDAE	VOTO - D
65_13011			TPTO PTTTTTTTUDKENT SKI EKV STRONDO STT
65_22/22	EFVREHDFDGV	DIDNE	TPRG-TDUILNYALLIKELRERFDEE-F
C04F8.3	ITFVRTWGFDGI	DIDNE	YPSG-ATDMANYVALVKELKAACESEAG
BUX \$00422 489	I E EVRKYG EDG I	DIDNE	PKG-DTOKONENLELGELKHACEOFAT
Mb100200708 Constat278 (500)	LAEVELUNENC	DIDUE	VOKO - F
annogrande Contograne 45201.		E E E	ADRUMPHITELNET NUMPHIE SA
GPLIN_000891100	I SEVRKYGEDGI	DIDNE	YPSG-ETEKQNFNKFLKELKDAIGAESY
GPLIN_000183600	* NG LDG 4	DIDWQ	FPTVNDKANYVSFLRELNMALPS
GPU IN 001228900	NGI DG	DID AD	FRTVN OKANYVAFIRFINMALPS
TIONS 2	IPELVOUELOC		
110110.2	* INFLKONELUGV	E L L V V V	EILTGELLUNLKMGLEKU
T19H5.3	* I SI LEEYQLDGV	DILLNK	VAKN S N TKKC SR FLCELKQKLKER
T19H5.6	LELLEKYDLDGV	DLFNR	VPK SD DKDEYAVFLRELKKQLKAR
C08H9 13	VEFIKSNNASGY	DIFAR	VON LSE
Door is the			
RUSD1.11a	* LEFLEATOLDGV	DILLINK	WPEEETRUATPAPINGLERKIPEAK
M01B2.6	* F S F L E E H K L D G V	DIFAK	VPHLADKHAYSQFLLELNEILK-T
R09D1 10	+ VSELKKYDIDGV	DIFAK	VPRIEDKYYYSRFISMLKERFOME
000017	TELOIVOIDOI	DI ENG	VOCCA
RUSU1.1	* I SP LUT TO I DOI		WPGGR SKILTIKFI SULKINLUUK
R09D1.6	*VSFLQTYDIDGV	DLYNK	WPGKT SKDIYSQFINDLRYSLQRQ
K08F9.3d	* T S F I EKYQ VDG V	ELHWN	INEHFLSQLETTRNLKNRLKKI
KOSER 32	TSFLEKYOVDGV	ELHAN	IN EHELSQLETTRNLKNRLKKI
KORED SD	TRELEKYOVOCV	e	
RUOF 8.30	*ISFIEKTQVDGV	E L B 3 3	INERFLOWLETTRNLKNRLKKI
R09D1.8	* VSFLQTYDIDGV	DILYNK	WPGKT SKDIYSQFINDLRYSLQRQ
T10D4.3	* SSFLAYHELHGV	DIFNK	VP SPE DKAHY SRFLADLRQHLG
C45E5 3	LALLIEHNIDGY	DIFAR	IN PAA
CICALO.			
F15A4.8C	*LSFLIERKIDGV	DIFNK	WPKLG
K08F9.3c	* T S F I EKYQ VDG V	ELHAN	INEHFLSQLETTRNLKNRLKKI
GPLIN 000538800	*		
F1544 82	-I SELLEHKIDGY	DIFUS	VOKIG
FISH4.08	E STETERKIDG V		WERE CONTRACTOR ENGLISHING TO THE CONTRACT OF THE CONTRACT.
F15A4.8D	* LSFLIEHKIDGV	DILWK	WPKLGNKASYSKFLKELNQKLKNQ
R09D1.5	* ARFIRQHKIHGV	DIYNK	VLGN SE TEHHD FP SFLKDLKEKLKTV
M176.8	LAFLVEHRIDGV	DFFND	VPTVQ DTFNYVTFIRELROKLDEN
P0901 3	- LAFTHSHMIDGY	DIHOK	VOT SR
00001.0			
R09D1.1	* VSFLIENGIDGV	DIFIK	VSSSRUKFNFSVFMRULREKLUKG
R09D1.2	* I Y FLKQ YKLDG V	DILFWK	VSSSGDKFTYSSFLQELKHKLRSH
C08H9.7	*VSFVKNEDIDGI	EIFWM	VPKHRD VNNYSIFIQDLRNEFTELQK
C08H9.11	+VSFFKKNDIDGI	EIFAT	RPKYED IKSYSSFIGELRSAFTFICK
COSHO (D	W SEVER VOIDO	ENCLOSE	NOR YRD
Coorte.40	*VSFVERTUIDGI		WPRTRD KARTEMPTWEERTAFTEEWA
C08H9.10	* VSFVQKYDIDGI	EINI	VPKYED ANNYLIFIQELRYAFIELQK
C08H9.12	* V S F VQN YD I DG V	DIFWT	VPKYSD ENNYLMFIRELRYAFTELQK
7K938 6	+ ATEVROYDIDGY	DIENT	VPKH SD ENNYLMFIRELRYAFTELOK
CE 08074	I STURRYHEDOL	C C V C D	TEON WEINPEEKELOLITEKTOEDAT
65 00974	*IEIVAATHPDGL	a a a a a a a a a a a a a a a a a a a	SEWHTTERTERTERTERTERTERTERTERTERT
Bm965	*		
Bm8301a	*		• • • • • • • • • • • • • • • • • • • •
Bm8301c	*		
Bm10790b			
Dini 107380	*		MATATVEEAK
Bm10739c	*		• • • • • • • • • • • • • • • • • • • •
Bm12106a	*		
Bm1951b	*		MKTAFVEFAK
DUV change the	1		THE REAL PROPERTY OF THE PROPERTY OF THE REAL PROPE
BUX 501038.115	*		
Mn10g200708_Cont/g914_68531_1	*		
Mh10g200708 Contlo343 40551	*		
Mb100200708 Constants (2011			
00 00500			
65_22506	*		
GS_04985	*		MLAAAEAEAA
GS 01873	*		
CPI IN 001120800	*		
GPLIN 001591000	** * * * * * * * * * * * *		
	*		
GPLIN_001586500	*		

**Figure S6.** Alignment of the catalytic site of GHF18 chitinases. Amino acids are represented by one letter codes in different colors. The three catalytic residues (DXDXE) are boxed in blue. Sequences that do not contain all three key residues are marked with an asterisk. Marked sequences may be considered as potential chitinase-like lectins. The top sequence is a reference human GHF18 chitinase (Q9BZP6.1).

Mh10g200708_Contig1101_12080	VPL	ΗV	NV	R	FDEGK~\	/ V	L -	N	S~AGE	WE K -	-	E	E	R -		HSN	- P F
gnl MincDB prot:Minc06838_1	TIF	HF	NP	R	F 1	/ V	C -	N	R~N-G	NGT -	-	E 1	٧ ١	RD	1	GGF	- PF
BUX.s01147.236 1	IAL	HF	NL	R	FAERS-1	ΓV	R -	N	A ~ - G N	NGG -	-	E	E	RD	1	GGF	- PF
Y55B1AR 1 1	IVI	HE	N	R	FDEGA	10	N.	N	S GG	NOS.		F	D	RH		- A N	. PF
GPI IN 000927200 2	1 41	U v	ND		EDOK	1	D		A	MGN			-	DE		G K M	PI
DUV -01100 011 0	1.25			12	EDEK .				OL	ALC N							
BUX.S01109.344_2	105		NP	5	FUER		R -	E.	A~ - NE	NGN -	-			RE		JAI	
ZK892.16_2	IAL	HF	NP	н	FDEK	/ V	R -	N	S~-GE	WGN -	-	1		RE		GKN	- P F
ZK892.1d_2	IAL	HF	NP	R	FDEKA	/ V	R -	N	S ~ - GE	NGN -	-	E E	E	RE		GKN	- P F
ZK892.1a_2	IAL	HF	NP	R	FDEKA~\	/ V	R -	N	S ~ - GE	NGN -	-	E I	E	RE	1	GKN	- PF
ZK892.1f 2	IAL	HF	NP	R	FDEKA	1	R -	N	S ~ - GE	NGN -	-	E E	E	RE	1	GKN	- PF
F52H3.7a 2	IAL	HL	N	R	FDEK	1	R -	N	S ~ - S A	NGN -	-	E F	E	RE		GKM	- PF
E52H3 7h 2	1.41	HI	N	R	FDFK	14	R.	2.1	4 2 2	MGN -			-	RE		GKM	PF
Pm/277b 2	111	02	NI D	12	EDEK	L.		2	C. CE	AGN				BE		CKH	VI
Dill42110_2	1.2.5			12	PDE K-			1	SOL	NON-						O K M	
65_23357_2	AL		NP	P.	FUEKS~	Y	<b>R</b> -	IN.	S~-GE	WGN -	-			RE		5 K N	- PL
GPLIN_000387400_2	IAI	HL	NP	F	FDEK	V	R -	N	S ~ - GE	MGN -	-	E I		RE	!	<b>SKI</b>	- PL
BUX.s01109.574_2	IAL	HF	NP	R	FDEK - ~	I V	R -	N	S ~ - NE	NGN -	-	E I	E	RE	!	GKN	- PL
gnl MincDB prot:Minc07760_2	IAL	HF	NP	R	FDEK -~	V	R -	N	S ~ - NE	NGN -	-	E I	E	RE	1	GKL	- PL
C44F1.3 2	ILF	HF	NP	R	IKDKA-V	/ V	R -	N	5 ~ - GF	NDT -	-	E I	E	RE	1	GGF	- PF
GS 21792 2	ILF	HF	NP	R	FKEKK-	1	R -	N	A ~ - G N	NGT -	-	E 1	E	RE	(	GPF	- PF
GPL IN 000653100 1	VVI	HV	NP	R	FKELE	v	1	N	N AG	NOK -		F		R		HRN	- KE
WOIAIIA 1	VVI	HV	N		FHHEH .	1	L.	N	S. GH	MGP		-		R		HKN	PI
002 22 1	KAL			2	ED AKT				C C C	MGH	-			2		E D H	P
002.38_1	NAL.	35	a P	2	C D A K T ~ L			1	S . OK	NON-	-			2.		UN	
DC2.3D_1	KAL	5 F	NP	R	PUAKT~L	V	1 -	N	5~ - GR	WGN -	-			R -		- UN	- PF
Bm3068d_1	KAL	HF	NP	R	FDTGS~I	V	L -	N	5~ - NR	WGM -	-	E		R -		YAN	- V F
GS_21591_1	KAL	HF	NP	R	FETSS~	V	L -	N	S ~ - N R	NGV-	-	E I	E	<b>R</b> -		YGN	- V F
GPLIN_000150600_1	KAL	HI	NP	R	FDSSG~	V	L -	N	S ~ - N T	NGV -	-	E (	Q	<b>R</b> -		LPN	- PL
gnl MincDB prot:Minc04346 1	KAL	H I	NP	R	FDRG I	V	1 -	N	S ~ - NV	NGT -	-	E (	Q	R -		NQN	- NL
anllMincDBlprot:Minc08713a 1	KAL	H I	NP	R	FDRG -~ I	v	1 -	N	S ~ - NV	NGT -	-	E (	0	R -		NQN	- NL
Mh10a200708 Contia349 18888 2	VAL	H	NP	R	FDRG	v	1 -	N	SNV	NGA -		F (	0	R -	1	NSN	- 51
GPLIN 000653000 1	VVI	HV	NP	R	FK	v	1 .	N	N ~ G	NOK -		F		R .	1	HRN	
GS 20599 1	VPI		c .		EDECK I	1			TNOF	MOK				2		K T I	PV
CS 19091 1	VPI		0	12	EDECK		-		C KG	MCK			-	2		OK I	P
7/002 4- 4	VDL	S. 1	0	12	FDEGK			1	KOT	MOK-				2.			
ZK892.18_1	VPL		5		FDEGK~	V	-	IN I	A~KGT	NGK -	-		E	R -		A K N	- 1
ZK892.16_1	VPL	HL	5	R	FDEGK~	V	۲ -	N	A~KGT	WGK -	-	E	E	R -		AKN	- P I
ZK892.1f_1	VPL	HL	S	R	FDEGK~	V	Υ-	N	A~KGT	MGK-	-	E	E	R -		AKN	- P I
ZK892.1c_1	VPL	ΗL	S	R	FDEGK~	V	Υ-	N	A~KGT	WG K -	-	E	E	R -		AKN	- P I
ZK892.1d_1	VPL	HL	S	R	FDEGK~	V	Υ-	N	A~KGT	NGK -	-	E	E	R -		AKN	- P I
Bm6267c_1	VPL	H I	S	R	FDEGK~	V	М -	N	S~NGE	NGK -	-	E	E	R -		KNL	- PF
BUX.s01109.344 1	VPL	H I	SV	R	FDEGK-	I V	L -	N	S~NGE	NGK -	-	E	E	R -		KSN	- P I
W09H1.6a 1	VPL	HV	s v	R	FDEGK~	v	L -	N	S~NGE	NGK -	-	E	E	R -		KSN	- P I
W09H1 65 1	VPL	HV	s v	R	FDEGK~	v		N	S-NGE	NGK -		F	F	R -		KSN	- PI
BUX \$01109 574 1	VPI		S V	R	EDEGK	Ń	E.	N	TAKOF	NGK.				R .		KGN	PE
E5242 7a 1	VPI		cu	12	EDECK I				TKOF	NGK			-	2		K C N	PV
E52H3.78_1	VPL	3.	0	12	EDECK-L	- 1			TKOE	MCK				2		K C M	
F52H3.7D_1	VPL	5 Y	2 4		PDEGK-I	1	5.1	5	TANGE	NON-	-			R -		N S IV	
Mh10g200708_Contig168_118666	VPL		5 V	Ь	FDEGK~	Y	<b>T</b> -	N	I~KGE	WGK -	-	E	E	R -		KSN	- P 1
gnl[MincDB]prot:Minc07760_1	VPL	H	SV	F	FDEGK~	V	Υ-	N	T~KGE	WGK-	-	E	E	R -		KSN	- P Y
gnl MincDB prot:Minc13870_1	VPL	н	SV	R	FDEGK~	V	Υ-	N	T~KGE	NGK -	-	E	E	R -		KSN	- P Y
GS_23357_1	VPL	H I	SV	R	FDEGK~	V	F -	N	T~KGE	NGK-	-	E	E	<b>R</b> -		KSN	- P Y
Bm11256_1	VPL	H I	S	R	FDEGK~	V	F -	N	T~KGE	NGK -	-	E	E	R -		KSN	- P Y
Bm4277b_1	VPL	H I	S	R	FDEGK~	V	F -	N	T~KGE	NGK -	-	E	E	R -	1	KSN	- P Y
GS 21792 1	VIL	HV	SV	R	FDEGK~	V	L -	N	S~RGT	NGK -	-	E	E	R -	1	VSN	- PF
GPLIN 000842700 1	API	HC	SV	R	FDEGK-	IV	1.	N	S-AGE	NGK -	-	E	F	R -	1	HSN	- PF
anIIMincDBlorot-Minc06831 1	TIE	HF	NP	R	F	1	R	N	T.T.G	NEK -		F		RY		GGE	.PF
anilMincDBioros-Mine(2/27_2	TIF	u e	NP	E	F	1	P	N	TTTO	MEN		F		RV		GOF	.PF
gnillMinoDPloret Minet2421_2	+			2	-		0	1	TT	MGM				24		GOF	P
gnillMineDBjprocivine12424_2				12					T. T. C	NOR-	-			2			
gniiwincDBiprot:winc12425_2			NP			11	R -	N	1~1-0	NGR-	-			R		30F	
W09H1.6a_2	ISF	HF	NP	F	FDEK -~ 1	1	R -	N	S ~ - NE	NGN -	-	E I		RE		GKN	- P F
W09H1.6b_2	ISF	HF	NP	R	FDEK - ~ 1	1	R -	N	S ~ - NE	NGN -	-	E I	E	RE	(	GKN	- P F
GS_19091_2	IAL	HF	NP	R	FDEK - ~ 1	1	R -	N	A ~ - GQ	NGK -	-	E I	E	RE	1	GKN	- PF
gnl MincDB prot:Minc13870_2	IAL	HF	NP	R	FDEK - ~		R -	N	S ~ - NE	NGN -	-	E I	E	RE	(	GKL	- PL
Bm2767b 2	LLF	HF	NP	R	FKEKQ	11	R-	N	A ~ - D 1	NGQ -	-	E I	E	RE	1	GIF	- PF
Bm2767d 2	LLE	HF	NP	R	FKEKO	1	R -	N	A D I	NGQ -		E	E	RF		GIF	- PF
Bm2767c 2	11.5	HF	NP	R	FKFK0-1	1	R	N	A	NGO.		F		RF		GIE	.PF
GPLIN 001119900 2		av	0		VSEG		D		T	MGD				20		GOF	PI
Bm7649a 2	LAL	S .	0	2	ENECT			1		MGD-	-					GN	
D-70104 0	TAL	5	2 4	12	ENEGI~I	V	K -	TV.	~ - ND	NOR-	-					JV I	
Bm/6480_2	HAL	9	5 V	5	NEGIA	V	K -	51	ND	WGR -	-			RE	1	JVI	- 1
Bm/648e_2	HAL	GI	SV	R	FNEGT~	V	R -	N	A~ - ND	NGR -	-	1		RE	(	311	- P I
BUX.s01109.342_2	YAL	GL	SI	R	YNEGV~	V	R -	N	A ~ - NE	MGK-	-	E I	E	RE	1	GGF	- PL
GS_06140_2	YAL	QV	S	R	YGEGA~	V	R -	N	A ~ - NV	NGT -	-	E I	E	RE	1	GGL	- P I
ZK1248.16_2	YAL	GV	SI	R	YNEGA~V	V	R -	N	A~ - NV	NGK -	-	E I	E	RE	1	GKM	-PL
and a second control of the second									1011 - 13000 Fe		- L		ante Te				

**Figure S7.** Alignment of the carbohydrate binding region of potential galectin sequences. Amino acids are represented by one letter codes in different colors. Only sequences containing all eight essential residues for carbohydrate binding are shown here. Key residues are boxed in red.

## Int. J. Mol. Sci. 2017, 18, 91; doi:10.3390/ijms18010091

41000				La L	P.C.		
1Kail	PCPWEWIFFQGNC YFM	SN SGRNVHD ST TACKE VG AGL VVTK SAE	Ean FLaLassk snk FlwmgLsDLnaEg IwawvDg	SPLLPSFKQTVM-RG	EPINN V	CEDCAE SGNGV	NODKCNLAKF-WICKK
gnl[MincDB[prot:Minc03866_1	RILNSFWLGMHRIQQYTSIC	• • • • • P • • • • • • • • • • • • • •			- 4 D R A V	TEDC TQLLD SA TGKV	NDIPCHYQQAGVVCKR
gnl[MincDB[proz:Minc10698_1	KMYRSLYIGLNRAAIEPYYE	YASVP	TPG VPTTIIPPWGAGC P\$GLNDTGNT		- PGKG	FEDCVCMFKYD EVNV	ND
Mh10g200708_Cont/g34_17139_20	SNLRVLVVLGMHR SAIYPYYN	NTLD \$1N \$DG TP - CD YFN I T	\$GTED\$		- T AN T Q	GEACVVMYSSTTGAV	ND
gnl[MincDB[prot:Minc09302 2	WR SLWLGLHRCAFYPTYN	ATVDCIN \$DGTV-CD YLNIT	P\$GTETG		- T 4 \$G Q	GEACAAMY SATTGQV	NDIACFNKL
anil MincDBI proz: Minc05278a 2	FSSLWVGLHRSAFYPSYN	YENIT	GG P \$G TETG		-G4\$GQ	GEACAAMY SAT TGQV	NDIACENK L
anil MincDBI proz-Minc05278d 2	FSSLWVGLHRSAFYPSYN	\$TVDCIN \$DGTI-CDYENIT			- G 4 \$G Q	CEACAAMY SAT TGQV	NOI ACENK
anii MincORI proz-Minc05278a 2	F S S I MA/C I HR SAFYD SYN	STUDCINSDOTI CD	CO			CEACAAMY SAT	NOI ACENK
grillanicoblproc.anicobl/ce 2						TO OUT	
grifmineDB[proc.mine138378_2		sivucinsuori -cu				UCAC AAMT SAT	ACTINK
gnilmincDB[prot:minc14997_1						SEDCIGITLAIS	ACHOKIGGFICKK
Mh10g200708_Cont/g34_17139_20	MHRCQYYPTYN	····· ATLDC VN ADG TP - CD ····· YAF ···	SKAVNTAGFQPYPWGWGC P SGTDGG - NG	•••••••••••••••••••••••••••••••••••••••	••••G•••••••••••••	VEDCIAMENASSGLV	NDY SCFQK MGGVVCKR
Mh10g200708_Contig1203_18954_	MTRKYIEET	DT FK \$EW\$DG T FV - K YGD VP	IRVAFHTPPVMGGQPDF	\$	G	VEECVH SYPKLGCDN WQQIFFNQV	NDARCLIK L SGAVCKK
GPLIN_001347500_1	MIGLRRVYAG	FGNVP	EITTVESTSGGRNKYPWGSGQ PSGTNKWTDP	T	G E	KEECVHMFSADGQV	NDAGCEIQMTGAVCQK
GPLIN 000610300 1			EITTAESTTGGRNTYPWGSGQ PSGTNKWTDP	T	G T	KEECVHMFSAD GQV	NDAGCEIQ MTGAICQK
GPLIN 000610400 1	MIGLRRVQAQAG	GQ YQ \$ TWANG T & A - T FGN VP	EITTAESTTGGRNTYPWG SGQ P SGTNKWTDP	T	G T	KEECVHMFSAD GQV	NDAGCEIQ MTGAICQK
GPL IN 001380700 1	MIGIRRVQAQAG	GOYO STWANGTAA -T FONVP	FITTAF STTGGRNTYPWG SGO P SGTNKWTDP	T		KEEC VHMESAD	NOAGCELQ MTGALCOK
Penticia 1	NI N SN AWL CHOM		DC RVPWI DO	DIVIIRC. DD	URVIRICODRAL FOUNDEDOF	DATIMNO VAL	NUL ENKTR
00 07751 1	DUNE NOR CLUK		CONTRACTOR OF CONTRACTOR	TINET	RAN TRIGODRATTON VOLPOTS		ACRYPT PORTAGE ACTION
GS_0//51_1	DENEMTWIGLAR		SORWEWEDG	LINTINVI - EG	EPRECCON	HACVIV	UDS SCRTRIRPTRM@STVCKK
GS_0/08/_1	DGEST WIGLAN K	• • • • • • • • • • • • • • • • • • • •	GG I WHWI DG	I PLIYIKWR - EG	EPDGCCAA	EAD TAAANYG SD TG EV	UDGPSVSSWH NTHNYVCKK
GS_03918_1	RGAYRYWIGLNKI	· · · · · · · · · · · · · · · · · · ·	NG TWRL \$NG	0 I PT F TNWR - I T	GPDLCCGL	NVTCVLVNYIDL SGLV	ODAGCNYIWT QPQGYVCKI
GS_17199_1	RNVRYYWIGLNR M		FGRWVWTDG	T P A L Y T NWR - PG	EFDGCCGA	NTTCALV NYGNTLGQV	DDAGCETLWA NPQGFVCKK
GS 20652 1	SSASPYWIGLNK M	<mark></mark>	FGQWWWSNG	M I AMF SNWR - PG	Q PD G C C A A	NVTCVLVNYAGNNGQV	DDAGCG \$ IVK N PQG FVCKK
GS 15951 4	LAVSSAFIGLVKQ -G			\$ PVTFVKWD - KG	EPDD SLGR	DN - EECV QMKLKEDFAV	GTVTCWQN RHFICSV
K02F3.5 1	NGTAVVIGVNDIGK		ENVERN SDG	TPVDFYKWG-KK	OFDNGEHN	EN-CVEVDH SGOV	TOKLCIIT RPEVCKK
GS 01335 1	KWF AWIGLNDIOK		ENKYWARDR	S SEREKNUM .KC	OFON AD YN	EN-CAEMG	DSECMAG RHYLCKR
GS 14289 1	- P SC FWI CMSRD - V		Y & PIAIEIAIT DE	TANK AND AND A DO	OCT NYNEH	FALCY YEN	FOX SCHOR SINSSICKE
CILARI 2	KYEHI AWIGHER		CONVERTICE.	D SD V ENAME OF	OCT UDVVE	NOTADOL S	NO SCHEV.
	K TENE AVVIONNK - V		COWWIDG	P SUTFRVVA - PK		NOT A POLATION NOT A POLATION	LEGHEV MRATICKK
W04E12.6_2	KTEHLAWIGMRKV	• • • • • • • • • • • • • • • • • • • •	GODWW/TDG	PSDYFNWA - PK	UPUNPKKE	CO-VOIAPDLSHDKWYENV	LECHEVMRAYICKK
W04E12.8_2	NPKOLAWIGMHK V	· · · · · · · · · · · · · · · · · · ·	GQDW/WTDG	T P \$N Y I NWA - PK	QFUNPGKE	NO-VETAPDLSHDKWYENV	ACSTE MRAYICKK
F40G9.10_1	SKTNYVMIGLIFE		NREWSWIDG	\$	EFNNMKHE	LW - TALVSDVN PKKGQMYTHV	NNVN-MDRQRAYICKR
B0218.6_2	SH SDDVFIGLVY Q		N SKWQWTDG	\$ A VN F L NWG - DG	EFNNMDKE	VV - T \$ L V AD PH EDKN TEN TRV	NNVA-QIDMRAFICKK
T03F1.10 2	DWNDD WVIGLVY Q		N SKWQWTDG	\$ V VN Y VNWG - DG	EFNNMNKE	WW-TALVADPHEGKN SEGTRV	NNVP-QDDQRAFLCKI
B0218.8 2	DWTGDVFIGLVFQ		N SKWOWT DG	SVVDYVNWG-DG	EFNNMNKE	WW-TSLVADPHQGKYSEASRV	NNVP-QDDQRAFLCKI
R08C7 6 1	-KKVHHWIGNBBBBDCPGFY	F	HNKIHEWINN	VATEDI PIEL	ORDHAMKP	BOFCOSIAYOM	DOD SCEV.
E10E2 7 1	KNDI EMMI CAORKMIC DDAGI	NENCOERT NEEDNOA	D CDVC ENDUC	A KNODDELC SC VONNAELD		NERCHVANNOK I T	
F10F2.7_1	KIND TE WILL AURKMOCPDAG	Rentroof ST. BRANDO		V-ARREPOPTO 30 TORVALELO			
F10F2.0_1	KIPNR WWIG AURRPECHIEG	VEGRIGGEDT	N SR V F TVL HG	V-ALNPPOPIENWADPN		NEKC V V MIG SEVVIN	
F33E2.3_1	NIY FWVDGQRNNS	·····	DG F EWSDG	-TIGKAALGID-IA	MUD GM A - G	GENCLTIKYSPQNVEKPL	NOV SCIHNG G SGV SCGY
F11D11.14_1	NYG VWI DGRRNCT	••••••••••••••••••••••••••••••••••••••	NVI YWTDG	F-TTTFDALVST-NA	GIPSTKF-KG-G	VEDCLTIVPG \$	NDL SCGNTQL I LG FVC
Y17D7B.8_1	KG FWI GGGRNC P	TVPC	·····\$\$FRWVDG	Y - TT \$ YD A L V P \$ - N A	AUTVSKNTD A-N	SENYLTVFASNGTPLTI	NDINQDFT\$L IVGFVCGY
F32E10.3 1	QRLVQIG	L Y A P EG		NE INKTLL	EPYDH SM E-G	RERCGLLNVEKRVL	DDVDCEST SPDHH AQRYICQR
C50E3.14 1	D SRWG VWI DGKHQ S SC PTTA	TGCD	LTKNFSFRDT	T - LH FK AG YK WA - TG	EPNGA G-R	KFCCAYVYIT PNVQYHGMV	DDTMCY
C29F3 4 1			VND P F VWL NG	V	- SNDNTFANDYDSLYD-L	TODCL SMDLG	NOUTCDA F TAY SCGK
H02K04 2 1	VNEVI IGAKRYCA	VEPCA	TTOPYEMONO		STNNAL ACDEMPEY	TODCO SMEKD	TODA
0005351	TO FLUID ALL DOCA		TTOPTEVANO			COOL SHELL	TAVACCK
62873.51	SUPERIGATESCATT		TIDPTWWW		- SHUNNFANDFIGLTD-G	SOUCE SMOLA	INDITCOALTATSCOK
H02K04.1_1	TNFVHIGAMRVCA	······································	VIDP YEWLNG		-SINNALANDEMPLYD-G	SOUC V SMELP	NUTICUALTAYSCGK
Y118A8A.3_1	G TD SG TWLG AKC S S		SGC TWTDG	N - TVG TQG MM FA - PG	EPNQLSYP	- PCLYIWAKIG - DT LKRYPYGNGYI	DDTRLTT - AM MSYACGK
Y26D4A.6_1	VPMSGMVLGAQRNG		SGFYWTDG	F - TTGTAGFT YG - PG	QPDNMNIDAR-V	FOTCLILITMAP - SYTG WWTRFKNGEL	DD TWCN \$ - P \$ D P
Y53H1A.3_1	KTH SG VWLG AQRTS		ATAFRWIDG	H - TTGTEAMV FG - PG	GFDNDHGAG R - G	FGRCLQLMALTP - SYWNIPGHLINLKSGQL	DDIWCNV
M162.1_1	GANTITWIGGIRKAKCYRP-			YG - VN	QFDLMS-L	RGWCLTLLVHAT - SDAGVVTPQFSLYNGNN	DDVFCDQ - \$L DTYVCGK
W09G10.6 1	SG S I WI G LKRREECKN VGI	R	SDN SFEWTOK	\$ - T T G L D G I D WD - G G	GFONARR-Y	SGQCATLTAS HQR SV IG YHVGRL	DUVGCEF
Y59H11AR.5 1	SG SIWIGAKRTTLCKASR	L	TLTS FOWTOG	S - A SG LDG L I WN - NN	GEDNNY N-R	TOQCVVLLAA RTPTV SDDLQWG ANRL	DDC
ZK39.7 1	SG SIVVIGGRRTQACRNSA	L	SLN SFOWTDG	S-TTGTAGLVWAN-TN	OFONKD S-R	SCOCLVLLAS RASSIGDKWTWYANRN	00/
F36F12.5 1	-R SGG WWVG ARRRPACISSG		KTN SFYWTON	S-ATGINGMLFT-NG	EPMNGGVK-L	DODCALLTVA STPVV FNKFRTGON	00
F38F12 6 1	- V SGG WWWGARRRPACISSO		KTN SE WAT DI	S-ATGINGML	FRINGAS	DODCALLTVA STPLV MNAFRTOON	00
W10C11 13 1	N SG SLWAG LORVEKCLAOK	SATC S.	ARTA FOYTOO	S. VIGTOGEV		KONCALFIAS RTOTI .VSC SEFACTI	
14/10/211 14 1	C SAME LOPTOKOL KOK	TARCE	ALT FEUTON	VICTOCEV.	old nu e	NONCALLEAS WITCH INCOMPANY	COT N
140001115 1	AC STANO NO PTOKOL KON	EDTC :	ALTETION	VICTOCEL CONTRACT	CONTRACT OF CONTRACT OF CONTRACT	NOR AVILLAS	
10011.15	- as a www.smartakclkapi		ALTAPETIOG	TTOTOG TOG TITTE GA		A A A A A A A A A A A A A A A A A A A	
W10G11.7_1	SA SVWIGIQRTKACVNKP		TN SFEWTON	S-IIGIDGFLFT-HE	GPUNH Y - 1	MOSCAILLAS KDPEIVLNGTYRAATL	P
W10G11.5_1	SG SVWLGLQRTARCMGQP	L	RTTAFEWION	S-ATGTDGFLFQ-TG	QPDNG R - L	NGNCALFLASIDPFIDARGRYYAATF	EDV
W10G11.6_1	SASVWVGIQRTAACTRQQ	I TAACT	RTT\$FEWTDG	S - ATGTDGFV FQ - PG	GPONI L - L	NGNCALLLAS RTPTITARGTYYAATL	EDVNC
W10G11.6_2	SASVWVGIQRTAACMNQP	P	RTTAFRWTDN	S - ATGTAGFV FQ - PG	QFDNIR-L	NGNCALLLAS ETPTITARGTYYAATL	EDVN
ZK39.5 1	SG S FWI G ARRRAACQ TOR	TADCT		T - VTG TDG FV WN - DK	Q F D N S D - L	NSGCAVLLAS GPPVQWGNGMWAGAKL	DDHV
ZK39.6 1	SG S FWIGAHRITTCMTOI	TAQCT	PTN SF SWTDN	S-ATRIDGEVWE-TO	G F D N S D - F	NSSCAVILAA NVPVMAAQRWWSPSKL	DDH
763931	SG SWALG ARR TD ACMTK S		ATN SETWIDG	\$ - T \$G T 4G FV WD - \$R	ORDND	KOPCV-ILLSSKTPFTPVSVKNRPVI PRMI	
763941	- SG SWUGAKRTAACTENG	\$ \$0CT	ATH SE VINT ON	S-ASGLAGEV	OCT NA	KTOLWALLESS RTSI RVKHVMAD DOM	
7690 2 1	SOCIMICARRITACARON	TTECE	TTAAEDUITAA	S. A SOTACEV.	OCON S.	NOCAVIHAC SSAATVSNAVADDA	DUTC
DODENING	AC ANNU O TRRTA AC ARUNI	NTNOCT	BTUSERWIDS		and the	TO SO AVELLAS BATUTION AVVID PAMA	
ROZENIN	36 3 VVI G I KK I A ACMKQVV	NTNOC	RINAF WIDG	S-AIGTAGEVWD-TL	K-L	SCAVELASRSIVINGGRPWQPAKN	
126E3.1_1	TG SVWIGAKRTTACLKQW	L · · · · · · · · · · · · · · · · · · ·	RQN SF AWT DG	5 - A T G V AG F V WD - N L	GPUND E - L	ROPCAVMLAA RAAVTWGGK FWQPAML	UDNNC
W09G10.5_1	TGTMVIGTKRTPPCMN \$G	V	QITSYYWTDG	\$ - T VG VQG F Y WN - \$G	EPNN-QGG-Q		DOVACTTELT NGYVCGK
W10G11.12_1	- PVGTLWVGAKRTKPCISSG	L		\$ - A VG VQG F F WR - AG	EFNNALG G - Q	- GCAQVYST TNDN	DDVGCGS-TNQGYICGK
Y51A2A_1_1	- PFDAIWIGAQRTTACLASV	Q	KDN SFEWTOM	\$ - T T G I D G F L W S - G A	DPNNGGP A - L	NEHCVIAFVD GPRL	ADISCA-RDFDGYVCGT
C31G12.2 1	- MD T T FWI GG ER T SACLT SGI	L	ALTSFEWTDG	\$ - TTG TDGWI WR - \$G	EFNNAGG L - E	I-CATVEITYVS-L	ADARCORADF NGYA
F16H6.2 1	- VEGAFWVGGQRTAACPGPG	L	ALTSFEWTDG	S-TTGTDGWIWR - PI	EFNNAAS	- CPVINAWT	ADQVCTR SDF DGYVCGK
F26D2 128 1	-GAGELWNGAORTSPELSSNI	DASCT	TINSFENTOO	KTTGTDGFL	OF DIGGA	DEACCLLSSA	SDEQCIRADP TO Y
F2602 120 1	GAGELWAYGAORT SPCI S SHI	DASCT	TINEFENTOC	KTTOTOCEL	OCONCO A CONCOLORIZACIÓN	DEACCI I SSA	SDEOCLEADE
F26D2 120 1	GAG FLWAG AORT SPCI 4 SHI	DASCT	TI N & E ENATED	KTTOTOCEL	OR NGG A	DEACCI I STA	SDEOCLERADE
FIGURA 1	TIOP DW COORTO LESSIN		I LN SFEWIDG			UTO CUTUUNI	
r48/10.1_1	- ILGP FWIGGURIGACISSG	IAIGI	ALN SF SWIDG	3 - A 10 1 AG P V	uruno GAM-L	GV-L	SUDACORTOG FGYACG-
		TATCT	RINEF SMITCO	A CAR BILL MARL CI	Manager of Cara and a second second second	ST WWENTY WAS A SAME	DC VACCV

**Figure S8.** Alignment of 77 potential carbohydrate-binding C-type lectin-like sequences. Amino acids are represented by one letter codes in different colors. Sequences were compared with a bacterial C-type lectin domain (1K9I). Key residues involved in carbohydrate binding are boxed in red. CTLDs were considered as potentially carbohydrate binding if at least three out of the five key residues were conserved.

<b>Table S1.</b> Presence of genes containing C-type lectin domains or hevein-like domains in different nematode genomes. This table gives an overview for several nematodes
of which the genome is known and indicates the number of C-type lectin domain and hevein-like domain-containing genes within each genome. Data were downloaded
from the Wormbase-Parasite website [3]. Nematode genomes used in the main analysis are in gray.

Species	Genome Size (Mb)	Life Style	Number of C-Type Lectin Domain Containing Genes	Number of Hevein-Like Domain Containing Genes	Phylogenetic Cluster (According to [4])	Provider	Reference
Acanthocheilonema viteae	77	Animal parasitic	14	0	8b	Blaxter laboratory at University of Edinburgh	
Ancylostoma ceylanicym	313	Animal parasitic	72	2	9b	Cornell University	[5]
Ancylostoma duodenale	333	Animal parasitic	76	3	9b	Mitreva laboratory at the Genome Institute of Washington University	
Ancylostoma caninum	466	Animal parasitic	81	2	9b	Mitreva laboratory at the Genome Institute of Washington University	[6]
Angiostrongylus costaricensis	263	Animal parasitic	8	0	9c	Parasite Genomic group at the Wellcome Trust Sanger Institute	
Angiostrongylus cantonensis	253	Animal parasitic	7	0	9c	Parasite Genomic group at the Wellcome Trust Sanger Institute	
Anisakis simplex	126	Animal parasitic	34	0	8b	Parasite Genomic group at the Wellcome Trust Sanger Institute	
Ascaris lumbricoides	317	Animal parasitic	29	0	8b	Parasite Genomic group at the Wellcome Trust Sanger Institute	
Brugia pahangi	91	Animal parasitic	13	0	8b	Parasite Genomic group at the Wellcome Trust Sanger Institute	
Brugia timori	65	Animal parasitic	10	0	8b	Parasite Genomic group at the Wellcome Trust Sanger Institute	
Caenorhabditis sinica	132	Free living	399	6	9a	Blaxter laboratory at University of Edinburgh	[7]
Caenorhabditis japonica	166	Free living	104	6	9a	Washington University	[7]
Caenorhabditis brenneri	190	Free living	264	5	9a	Washington University	[7]
Caenorhabditis tropicalis	79	Free living	221	5	9a	Washington University	[7]
Caenorhabditis angaria	106	Free living	261	3	9a	California Institute of Technology	[8]
Caenorhabditis briggsae	108	Free living	160	3	9a	Sanger Institute	[9]
Caenorhabditis remanei	145	Free living	235	2	9a	Washington University	[7]
Cylicostephanus goldi	173	Animal parasitic	30	0	х	Parasite Genomic group at the Wellcome Trust Sanger Institute	
Dictyocaulus viviparus	169	Animal parasitic	7	1	9B	Blaxter laboratory at University of Edinburgh	[10]
Dirofilaria immitis	88	Animal parasitic	20	0	8b	Blaxter laboratory at University of Edinburgh	[11]
Dracunculus medinensis	104	Animal parasitic	23	0	8b	Parasite Genomic group at the Wellcome Trust Sanger Institute	
Elaeophora elaphi	83	Animal parasitic	13	0	8b	Parasite Genomic group at the Wellcome Trust Sanger Institute	

Size

185

42

172

177

47

323

218

80

86

Animal parasitic

Animal parasitic

Free living

Free living

Free living

Insect parasitic

Animal parasitic

Insect parasitic

Insect parasitic

Species

Enterobius vermicularis

Gongylonema pulchrum

Haemonchus contortus

Haemonchus placei

Heligmosomoides polygyrus

Heterorhabditis bacteriophora Litomosoides sigmodontis

Loa loa

Necator americanus

Nippostrongylus brasiliensis

Oesophagostomum dentatum

Onchocerca volvulus

Onchocerca ochengi

Onchocerca flexuosa

Panagrellus redivivus

Parascaris equorum

Parastrongyloides trichosuri

Pristionchus pacificus

Pristionchus exspectatus

Rhabditophanes kr3021

Romanomermis culicivorax

Soboliphyme baturini

Steinernema scapterisci

Steinernema carpocapsae

			Table	<b>51.</b> Com.		
Genome Size (Mb)	Life Style	Number of C-Type Lectin Domain Containing Genes	Number of Hevein-Like Domain Containing Genes	Phylogenetic Cluster (According to [4])	Provider	Reference
150	Animal parasitic	80	1	8a	Parasite Genomic group at the Wellcome Trust Sanger Institute	
322	Animal parasitic	24	0	8b	Parasite Genomic group at the Wellcome Trust Sanger Institute	
370	Animal parasitic	87	1	9b	Parasite Genomic group at the Wellcome Trust Sanger Institute	[12]
259	Animal parasitic	34	1	9b	Parasite Genomic group at the Wellcome Trust Sanger Institute	
561	Animal parasitic	36	0	9b	Parasite Genomic group at the Wellcome Trust Sanger Institute	
77	Insect parasitic	15	1	9b	Mitreva laboratory at the Genome Institute of Washington University	[13]
65	Animal parasitic	17	0	8b	Blaxter laboratory at University of Edinburgh	
96	Animal parasitic	15	0	8b	Institute for Genome Sciences at the University of Maryland	[14]
244	Animal parasitic	49	1	9b	Mitreva laboratory at the Genome Institute of Washington University	[15]
294	Animal parasitic	30	1	9b	Parasite Genomic group at the Wellcome Trust Sanger Institute	
490	Animal parasitic	76	1	9b	Mitreva laboratory at the Genome Institute of Washington University	
96	Animal parasitic	17	0	8b	Parasite Genomic group at the Wellcome Trust Sanger Institute	[7]
112	Animal parasitic	13	0	8b	Parasite Genomic group at the Wellcome Trust Sanger Institute	
86	Animal parasitic	10	0	8b	Parasite Genomic group at the Wellcome Trust Sanger Institute	
65	Free living	134	1	10b	California Institute of Technology	[16]

Parasite Genomic group at the Wellcome Trust Sanger Institute

Parasite Genomic group at the Wellcome Trust Sanger Institute

Max-Planck Institute for Developmental Biology

Max-Planck Institute for Developmental Biology

Parasite Genomic group at the Wellcome Trust Sanger Institute

University of Cologne

Parasite Genomic group at the Wellcome Trust Sanger Institute

California Institute of Technology

California Institute of Technology

Table S1 Cont

10

29

151

116

34

25

4

295

187

0

1

2

1

1

0

0

2

2

8b

10b

9a

9a

10b

2a

2a

10a

10a

[17]

[18]

[19]

[19]

	Tabl	le S1.	Cont.
--	------	--------	-------

Species	Genome Size (Mb)	Life Style	Number of C-Type Lectin Domain Containing Genes	Number of Hevein-Like Domain Containing Genes	Phylogenetic Cluster (According to [4])	Provider	Reference
Steinernema feltiae	83	Insect parasitic	241	1	10a	California Institute of Technology	[19]
Steinernema monticolum	89	Insect parasitic	291	0	10a	California Institute of Technology	[19]
Steinernema glaseri	93	Insect parasitic	162	0	10a	California Institute of Technology	[19]
Strongyloides venezuelensis	52	Animal parasitic	28	1	10b	Parasite Genomic group at the Wellcome Trust Sanger Institute	[20]
Strongyloides ratti	43	Animal parasitic	27	1	10b	Parasite Genomic group at the Wellcome Trust Sanger Institute	[20]
Strongyloides stercoralis	43	Animal parasitic	27	1	10b	Parasite Genomic group at the Wellcome Trust Sanger Institute	[20]
Strongyloides papillosus	60	Animal parasitic	31	0	10b	Parasite Genomic group at the Wellcome Trust Sanger Institute	[20]
Strongylus vulgaris	291	Animal parasitic	32	1	9b	Parasite Genomic group at the Wellcome Trust Sanger Institute	
Syphacia muris	99	Animal parasitic	52	1	8a	Parasite Genomic group at the Wellcome Trust Sanger Institute	
Teladorsagia circumcincta	700	Animal parasitic	93	2	9b	Mitreva laboratory at the Genome Institute of Washington University	
Thelazia callipaeda	75	Animal parasitic	11	0	8b	Parasite Genomic group at the Wellcome Trust Sanger Institute	
Toxocara canis	300	Animal parasitic	27	0	8b	Parasite Genomic group at the Wellcome Trust Sanger Institute	
Trichinella spiralis	66	Animal parasitic	9	0	2a	Mitreva laboratory at the Genome Institute of Washington University	[21]
Trichinella nativa	49	Animal parasitic	7	0	2a	Mitreva laboratory at the Genome Institute of Washington University	
Trichuris muris	84	Animal parasitic	6	0	2a	Parasite Genomic group at the Wellcome Trust Sanger Institute	[22]
Trichuris suis	64	Animal parasitic	5	0	2a	Mitreva laboratory at the Genome Institute of Washington University	
Wuchereria bancrofti	77	Animal parasitic	17	0	8b	Parasite Genomic group at the Wellcome Trust Sanger Institute	
Caenorhabditis elegans	100	Free living	252	3	9a		[23]
Brugia malayi	88	Animal parasitic	18	0	8b	TIGR Institute	[24]
Ascaris suum	266	Animal parasitic	36	0	8b	Davis laboratory at the University of Colorado	[25]
Globodera pallida	124	Plant parasitic	22	0	12b	Parasite Genomic group at the Wellcome Trust Sanger Institute	[26]
Meloidogyne hapla	53	Plant parasitic	29	0	12b	Plant Nematode Genomics group at North Carolina State University	[27]
Meloidogyne incognita	86	Plant parasitic	57	0	12b	French National Institute for Agricultural Research (INRA)	[28]
Bursaphelenchus xylophilus	75	Plant parasitic	15	0	10b	Parasite Genomic group at the Wellcome Trust Sanger Institute	[29]

## References

- Gerken, T.A.; Revoredo, L.; Thome, J.J.; Tabak, L.A.; Vester-Christensen, M.B.; Clausen, H.; Gahlay, G.K.; Jarvis, D.L.; Johnson, R.W.; Moniz, H.A.; et al. The lectin domain of the polypeptide GalNAc transferase family of glycosyltransferases (ppGalNAc Ts) acts as a switch directing glycopeptide substrate glycosylation in an N- or C-terminal direction, further controlling mucin type *O*-glycosylation. *J. Biol. Chem.* **2013**, *288*, 19900–19914.
- 2. Ohnuma, T.; Onaga, S.; Murata, K.; Taira, T.; Katoh, E. LysM domains from *Pteris ryukyuensis* chitinase-A: A stability study and characterization of the chitin-binding site. *J. Biol. Chem.* **2008**, *283*, 5178–5187.
- Nicol, P.; Gill, R.; Fosu-Nyarko, J.; Jones, M.G. De novo analysis and functional classification of the transcriptome of the root lesion nematode, *Pratylenchus thornei*, after 454 GS FLX sequencing. *Int. J. Parasitol.* 2012, 42, 225–237.
- Van Megen, H.; van den Elsen, S.; Holterman, M.; Karssen, G.; Mooyman, P.; Bongers, T.; Holovachov, O.; Bakker, J.; Helder, J. A phylogenetic tree of nematodes based on about 1200 full-length small subunit ribosomal DNA sequences. *Nematology* 2009, *11*, 927–950.
- Schwarz, E.M.; Hu, Y.; Antoshechkin, I.; Miller, M.M.; Sternberg, P.W.; Aroian, R.V. The genome and transcriptome of the zoonotic hookworm *Ancylostoma ceylanicum* identify infection-specific gene families. *Nat. Genet.* 2015, 47, 416–422.
- Abubucker, S.; Martin, J.; Yin, Y.; Fulton, L.; Yang, S.P.; Hallsworth-Pepin, K.; Johnston, J.S.; Hawdon, J.; McCarter, J.P.; Wilson, R.K.; et al. The canine hookworm genome: Analysis and classification of *Ancylostoma caninum* survey sequences. *Mol. Biochem. Parasitol.* 2008, 157, 187–192.
- 7. Howe, K.L.; Bolt, B.J.; Cain, S.; Chan, J.; Chen, W.J.; Davis, P.; Done, J.; Down, T.; Gao, S.; Grove, C.; et al. WormBase 2016: Expanding to enable helminth genomic research. *Nucleic Acids Res.* **2016**, *44*, D774–D780.
- 8. Mortazavi, A.; Schwarz, E.M.; Williams, B.; Schaeffer, L.; Antoshechkin, I.; Wold, B.J.; Sternberg, P.W. Scaffolding a *Caenorhabditis* nematode genome with RNA-seq. *Genome Res.* **2010**, *20*, 1740–1747.
- 9. Stein, L.D.; Bao, Z.; Blasiar, D.; Blumenthal, T.; Brent, M.R.; Chen, N.; Chinwalla, A.; Clarke, L.; Clee, C.; Coghlan, A.; et al. The genome sequence of *Caenorhabditis briggsae*: A platform for comparative genomics. *PLoS Biol.* **2003**, *1*, e45.
- 10. Koutsovoulos, G.; Makepeace, B.; Tanya, V.N.; Blaxter, M. Palaeosymbiosis revealed by genomic fossils of *Wolbachia* in a strongyloidean nematode. *PLoS Genet*. **2014**, *10*, e1004397.
- Godel, C.; Kumar, S.; Koutsovoulos, G.; Ludin, P.; Nilsson, D.; Comandatore, F.; Wrobel, N.; Thompson, M.; Schmid, C.D.; Goto, S.; et al. The genome of the heartworm, *Dirofilaria immitis*, reveals drug and vaccine targets. *FASEB J.* 2012, *26*, 4650–4661.
- Laing, R.; Kikuchi, T.; Martinelli, A.; Tsai, I.J.; Beech, R.N.; Redman, E.; Holroyd, N.; Bartley, D.J.; Beasley, H.; Britton, C.; et al. The genome and transcriptome of *Haemonchus contortus*, a key model parasite for drug and vaccine discovery. *Genome Biol.* 2013, *14*, R88, doi:10.1186/gb-2013-14-8-r88.
- 13. Bai, X.; Adams, B.J.; Ciche, T.A.; Clifton, S.; Gaugler, R.; Kim, K.S.; Spieth, J.; Sternberg, P.W.; Wilson, R.K.; Grewal, P.S. A lover and a fighter: The genome sequence of an entomopathogenic nematode *Heterorhabditis bacteriophora. PLoS ONE* **2013**, *8*, e69618.
- 14. Tallon, L.J.; Liu, X.; Bennuru, S.; Chibucos, M.C.; Godinez, A.; Ott, S.; Zhao, X.; Sadzewicz, L.; Fraser, C.M.; Nutman, T.B.; et al. Single molecule sequencing and genome assembly of a clinical specimen of *Loa loa*, the causative agent of loiasis. *BMC Genom.* **2014**, *15*, 788, doi:10.1186/1471-2164-15-788.
- Tang, Y.T.; Gao, X.; Rosa, B.A.; Abubucker, S.; Hallsworth-Pepin, K.; Martin, J.; Tyagi, R.; Heizer, E.; Zhang, X.; Bhonagiri-Palsikar, V.; et al. Genome of the human hookworm *Necator americanus*. *Nat. Genet.* 2014, 46, 261–269.
- 16. Srinivasan, J.; Dillman, A.R.; Macchietto, M.G.; Heikkinen, L.; Lakso, M.; Fracchia, K.M.; Antoshechkin, I.; Mortazavi, A.; Wong, G.; Sternberg, P.W. The draft genome and transcriptome of *Panagrellus redivivus* are shaped by the harsh demands of a free-living lifestyle. *Genetics* **2013**, *193*, 1279–1295.
- Dieterich, C.; Clifton, S.W.; Schuster, L.N.; Chinwalla, A.; Delehaunty, K.; Dinkelacker, I.; Fulton, L.; Fulton, R.; Godfrey, J.; Minx, P.; et al. The *Pristionchus pacificus* genome provides a unique perspective on nematode lifestyle and parasitism. *Nat. Genet.* 2008, *40*, 1193–1198.

- Schiffer, P.H.; Kroiher, M.; Kraus, C.; Koutsovoulos, G.D.; Kumar, S.; Camps, J.I.; Nsah, N.A.; Stappert, D.; Morris, K.; Heger, P.; et al. The genome of *Romanomermis culicivorax*: Revealing fundamental changes in the core developmental genetic toolkit in Nematoda. *BMC Genom.* 2013, 14, 923, doi:10.1186/1471-2164-14-923.
- Dillman, A.R.; Macchietto, M.; Porter, C.F.; Rogers, A.; Williams, B.; Antoshechkin, I.; Lee, M.M.; Goodwin, Z.; Lu, X.; Lewis, E.E.; et al. Comparative genomics of *Steinernema* reveals deeply conserved gene regulatory networks. *Genome Biol.* 2015, *16*, 200, doi:10.1186/s13059-015-0746-6.
- Hunt, V.L.; Tsai, I.J.; Coghlan, A.; Reid, A.J.; Holroyd, N.; Foth, B.J.; Tracey, A.; Cotton, J.A.; Stanley, E.J.; Beasley, H.; et al. The genomic basis of parasitism in the *Strongyloides* clade of nematodes. *Nat. Genet.* 2016, 48, 299–307.
- 21. Mitreva, M.; Jasmer, D.P.; Zarlenga, D.S.; Wang, Z.; Abubucker, S.; Martin, J.; Taylor, C.M.; Yin, Y.; Fulton, L.; Minx, P.; et al. The draft genome of the parasitic nematode *Trichinella spiralis*. *Nat. Genet.* **2011**, 43, 228–235.
- 22. Foth, B.J.; Tsai, I.J.; Reid, A.J.; Bancroft, A.J.; Nichol, S.; Tracey, A.; Holroyd, N.; Cotton, J.A.; Stanley, E.J.; Zarowiecki, M.; et al. Whipworm genome and dual-species transcriptome analyses provide molecular insights into an intimate host-parasite interaction. *Nat. Genet.* **2014**, *46*, 693–700.
- 23. Sequencing Consortium. Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science* **1998**, *282*, 2012–2018.
- 24. Ghedin, E.; Wang, S.; Spiro, D.; Caler, E.; Zhao, Q.; Crabtree, J.; Allen, J.E.; Delcher, A.L.; Guiliano, D.B.; Miranda-Saavedra, D.; et al. Draft genome of the filarial nematode parasite *Brugia malayi*. *Science* **2007**, *317*, 1756–1760.
- 25. Jex, A.R.; Liu, S.; Li, B.; Young, N.D.; Hall, R.S.; Li, Y.; Yang, L.; Zeng, N.; Xu, X.; Xiong, Z.; et al. *Ascaris suum* draft genome. *Nature* **2011**, *479*, 529–533.
- 26. Cotton, J.A.; Lilley, C.J.; Jones, L.M.; Kikuchi, T.; Reid, A.J.; Thorpe, P.; Tsai, I.J.; Beasley, H.; Blok, V.; Cock, P.J.; et al. The genome and life-stage specific transcriptomes of *Globodera pallida* elucidate key aspects of plant parasitism by a cyst nematode. *Genome Biol.* **2014**, *15*, R43, doi:10.1186/gb-2014-15-3-r43.
- Opperman, C.H.; Bird, D.M.; Williamson, V.M.; Rokhsar, D.S.; Burke, M.; Cohn, J.; Cromer, J.; Diener, S.; Gajan, J.; Graham, S.; et al. Sequence and genetic map of *Meloidogyne hapla*: A compact nematode genome for plant parasitism. *Proc. Natl. Acad. Sci. USA* 2008, *105*, 14802–14807.
- Abad, P.; Gouzy, J.; Aury, J.M.; Castagnone-Sereno, P.; Danchin, E.G.; Deleury, E.; Perfus-Barbeoch, L.; Anthouard, V.; Artiguenave, F.; Blok, V.C.; et al. Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita. Nat. Biotechnol.* 2008, 26, 909–915.
- 29. Kikuchi, T.; Cotton, J.A.; Dalzell, J.J.; Hasegawa, K.; Kanzaki, N.; McVeigh, P.; Takanashi, T.; Tsai, I.J.; Assefa, S.A.; Cock, P.J.; et al. Genomic insights into the origin of parasitism in the emerging plant pathogen *Bursaphelenchus xylophilus*. *PLoS Pathog*. **2011**, *7*, e1002219.