Does Oat Have a Grain Hardness Locus?

Michael A. Tanchak[1]*, Martin J. MacInnis[1,2], Kelcie A. Lahey[1,3], Alicia M. Currie[1], Charlene P. Wight[4] and Stephen J. Molnar[4]. [1]Dept. of Biology, Cape Breton University, P.O. Box 5300, Sydney, NS B1P 6L2, CANADA. [2]Current Address: School of Kinesiology, University of British Columbia, 210-6081 University Boulevard, Vancouver, BC V6T 1Z1, CANADA. [3]Current Address: Dept. of Physiology and Pharmacology, University of Calgary, 3310 Hospital Drive NW, Calgary, AB T2N 4N1, CANADA. [4]Eastern Cereal and Oilseed Research Centre, Agriculture and Agri-Food Canada, Central Experimental Farm, 960 Carling Ave., Ottawa, ON K1A 0C6, CANADA.
*Correspondence: michael_tanchak@cbu.ca; presented by: michael_tanchak@cbu.ca

Introduction: The *Ha*-locus or Hardness locus of wheat [*Triticum aestivum* L.] is located on the short arm of chromosome 5D and contains genes that encode for puroindoline a [Pina], puroindoline b [Pinb] and grain softness protein 1 [Gsp1] [reviewed in Ref. 1]. This locus is of interest because a substantial body of evidence indicates that the puroindoline proteins [Pins] are the major determinants of grain texture/hardness, a major agronomic and quality trait, in wheat [reviewed in Refs. 2 and 3]. In addition, all three proteins and/or synthetic peptides derived from them have demonstrated antimicrobial properties [e.g. Refs. 4-9]. Pins have a distinctive "tryptophan-rich domain" [TRD], amino acid sequence WRWWKWWK in Pina and WPTKWWK in Pinb [Ref. 10] , which influences their lipid-binding properties, antimicrobial activity and effect on grain texture. The Pin genes are missing from the corresponding regions of the homeologous wheat chromosomes, chromosomes 5A and 5B [e.g. Refs. 11-12]. Comparable *Ha*-loci are found on chromosome 5H of barley [*Hordeum vulgare*] [Refs. 13-14] and on chromosome 5R of rye [*Secale cereale*] [e.g. Ref. 15]. Limited sequence information is available for oat [*Avena sativa*] but sequences showing homology to the Pins have been identified and are known as oat tryptophanins [OTs], 3B3 [OT3B3] and 3B3T [OT3B3T] [Ref. 16], or vromindolines [Vroms].

Methods: Putative oat homologues to *Ha*-locus sequences were identified as significant matches with the BLASTN program [Ref. 17] using the published Umn EST sequences in the NCBI [National Center for Biological Information, Bethesda, Md] EST database as query sequences of the NUCLEOTIDE database. Similarly, full-length ESTs for the relevant Umn sequences were identified using the relevant Umn sequences as query sequences of the EST database, but with the search restricted by species to *Avena sativa*. Amino acid sequences and additional sequence alignments were generated using the BioEdit program [Version 7.2.5, Ibis Biosciences, Carlsbad, CA]. OT154 sequences were identified using the phylogenetic analysis function of the BLASTN program.

DNA polymorphisms for OT3B3T and OT154 were identified by comparing the sequencing results of PCR products generated with sequence specific-primers and template DNA from various oat cultivars [Kanota, Ogle, Terra, Marion, Dal and Exeter]. Sequence polymorphisms generating differences in restriction maps were converted into CAPS markers.

Mapping data was generated by using the same primers and DNA from the recombinant inbred lines [RILs] from Kanota x Ogle , Terra x Marion and Dal x Exeter mapping populations. PCR products were digested with the relevant restriction endonuclease and subjected to agarose gel electrophoresis. Loci were mapped essentially as per Hizbai *et al*. [Ref. 18].

Results: Markers for the oat homologues to the *Ha*-locus genes were identified based upon results of BLASTN searches [Table 1]. These searches identified Umn287, Umn360, and Umn856 as oat homologues to Pina, Pinb and Gsp1. Umn287 corresponds to OT3B3/Vrom2 while Umn360 corresponds to Vrom1. There did not appear to be an Umn clone corresponding to OT3B3T/Vrom3.

Umn162, Umn249 and Umn753 were also identified as potential *Ha*-locus markers based upon alignment with the "hardness locus region" in *Brachypodium sylvaticum* [Table 1]. These ESTs show much more limited sequence identity with *Ha*-locus homologues from other cereals. This identity is restricted to the 5' and 3' ends of the ESTs [data not shown]. Umn162, Umn249 and Umn753 appear to represent three independent isolates of the same sequence as they differ only in overall length, in a small number of "uncalled" ["N"] nucleotides, and in minor differences in sequence at the extreme 3' ends.

None of the Umn ESTs identified in Table 1 appear to include a complete coding sequence. The relevant Umn ESTs were used to query the NCBI EST database to identify full-length ESTs. This process identified a variety of such ESTs from a set of seed-specific oat ESTs generated at the University of Saskatchewan [Ref. 19] [Table 2]. Unexpectedly, this process also identified a series of ESTs that appear to encode for a novel oat tryptophanin, OT154, predicted to be 154 amino acids in length and possessing a novel TRD [Figure 1, Table 2]. Sequences, represented here by accessions GO581309, GO581838 and GO582063, encoding for three variants of OT154 were identified. The TRD for OT154 is considerably longer than those of other OTs and includes a WKW [sometimes LKW] motif that is present as part of three copies of a 6 amino acid imperfect repeat [Figure 1, Table 2]. Exemplars of EST sequences containing full-length coding sequences for the various OTs are identified and partially characterized in Table 2.

With the oat *Ha*-locus homologues identified, their known loci in established oat genetic maps, Kanota x Ogle (KxO)[Ref. 20] and Terra x Marion (TxM) [Ref. 21], were reviewed. The strongest evidence of clustering, comparable to the *Ha*-locus in wheat, was seen in the KxO map where loci Umn287, Umn360a and Umn856a were placed on the same interval in linkage group [LG] 22_44+18 and additional loci, Umn360brv and Umn856b, mapped within 1 cM of one another on KxO LG 29_43 [Table 3]. Umn162, Umn249 and Umn753 co-localized on KxO LG 6 as would be expected if they were markers for a common sequence [Table 3]. In the TxM map, Umn360a and Umn856a were placed on the same interval in LG 15 [Table 3].

In an attempt to obtain additional evidence of marker clustering, PCR products corresponding to OT3B3T and OT154 were generated using the parental DNAs for the KxO, TxM, and DxE maps. Parental sequence polymorphisms were exploited to generate cleaved amplified polymorphic sequence [CAPS] markers for each of OT3B3T, OT154[GO582063], and OT154[GO581838] [Table 3]. See Table 3 for mapping outcomes.

	Key matches in NCBI NUCLEOTIDE Database						
Oat Marker [NCBI Accession Number]	Accession No.	Species	Sequence Type	Max Score	% Coverage	% Identity	E value
Umn287 [CK780259.1]	EF602433.1	*Avena sativa*	OT3B3	921	100	99	0.0
	JQ518370.1	*A. sativa*	Vrom2.2	782	83	100	0.0
	GU591260.1	*Hordeum vulgare*	Hordoindoline a	217	98	70	8e-53
	X69914.1	*Triticum aestivum*	Pina	212	98	70	4e-54
Umn360 [CK780266.1]	JQ518366.1	*A. sativa*	Vrom1.1	643	78	100	0.0
	JQ518372.1	*A. sativa*	Vrom3.1	148	65	76	8e-34
	AB446468.1	*Hordeum chilense*	Hordoindoline b	246	95	72	1e-61
	AY644058.1	*H. vulgare*	Hordoindoline b1/b2	232	95	72	3e-57
	DQ363913.1	*T. aestivum*	Pinb	221	96	72	6e-57
Umn856 [CK780284.1]	FJ898188.1	*Aegilops longissima*	Gsp1	307	97	72	7e-80
	EU307540.1	*Triticum urartu*	Gsp1	307	97	72	7e-80
	X80379.1	*T. aestivum*	Gsp1	306	97	72	2e-79
Umn162 [CK780288.1]	FJ234838.1	*Brachypodium sylvaticum*	BAC37D5, *Ha*- locus	211	88	79	6e-51
	FJ898198.1	*Aegilops comosa*	Gsp1	73.4	20	74	3e-09
	JN636843.1	*H. vulgare*	Hordoindoline b1	64.4	8	87	1e-06
	GQ496618.1	*T. aestivum*	Pinb-like protein	68	9	86	1e-07

Table 1: Identification of markers for putative oat homologues to genes of the *Ha*-locus. Identification based on BLASTN sequence alignments obtained using specific Umn marker sequences as query sequences.

Umn marker or oat tryptophanin sequence	Accession No. of "full-length" EST exemplar	Predicted size of polypeptide in aa	TRD present	Amino acid sequence of TRD region
OT3B3T	GO582128	142	Yes	PITWPWKWWKGGC
Umn287	GO584809	147	Yes	PITWPWKWWKGGC
Umn360	GO581793	147	Yes	PLTWPWKWWKGGC
Umn856	GO583768	163	Yes	PMIIWPWKWGKRSC
	GO584403	163	Yes	PMITWPWKWGKSSC
Umn162/249/753	GO585541	164	?	PGKMPFKWYKSC
	GO583453	164	?	PGKMPMKWYRSC
OT154	GO581309	154	Yes	PITQPWKWRKPWKWTEPWKWGMSSC
	GO581838	154	Yes	PITQPWKWSQPWKWTEPLKWGMSSC
	GO582063	154	Yes	PITQPWKWSKPWKWTEPLKWGMSSC

Table 2: Identification and partial characterization of exemplars for full-length ESTs of putative oat *Ha* homologues.

Map	Marker [Locus]	Linkage Group	Placement [cM]	Comment	Position of polymorphic site in EST	Polymorphism [Parent]	RE site [Parent]
KxO	Umn287[Umn287]	22_44+18	158-167				
	Umn360[Umn360a]	22_44+18	158-167				
	Umn856[Umn856a]	22_44+18	158-167				
	OT154(GO582063]	22_44+18	158-167	Co-dominant marker	Nuc. 329 in GO582063	T[T,E]/C[M,D]	*Sal* I [O]
	Umn360[Umn360brv]	29_43	0	Framework marker			
	Umn856[Umn856b]	29_43	1	Framework marker			
	Umn162[Umn162]	6	95-97				
	Umn249[Umn249]	6	95-97				
	Umn753[Umn753]	6	95-97				
TxM	Umn360[Umn360a]	15	13-18				
	Umn856[Umn856a]	15	13-18				
	OT154[GO582063]	15	13-18	Co-dominant marker	Nuc. 329 in GO582063	T[T]/C[M]	*Sal* I [M]
	Umn360[Umn360x]	????	????	Previously unlinked			
	OT154[GO581838]	????	?????	~7 cM from Umn360x	Nuc. 113 in GO581838	C[M]/T[T]	*Taq* I [T]
DxE	OT3B3T[GO582128]	???	?????	~3 cM from 4 GBS markers	Nuc. 273/4 in GO582128	AG[T,M]/GA[K,O]	*Msp* I [D]
	OT154[GO582063]	31	???	Co-dominant Marker ~ 7cM from oPt-794207 and oPt7947879	Nuc. 329 in GO582063	C[D]/T[E]	*Sal* I [D]

Table 3: Mapping data for relevant markers in three major oat genetic maps. Data in black type represent data previously published from other sources. Data in red type represent new mapping data reported here.

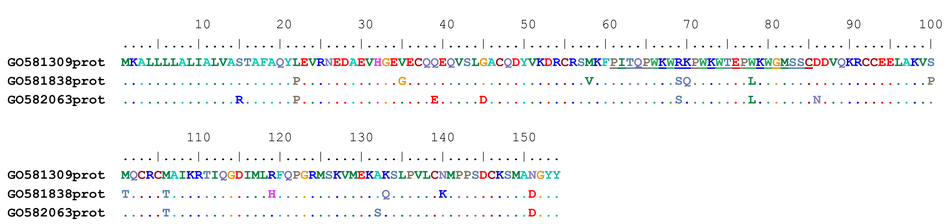

Figure 1: Exemplars for three variants of OT154, predicted amino acid sequences. Full amino acid sequence shown for GO581309, variant positions only for GO581838 and GO582063. Underlined amino acid sequence in GO581309 corresponds to the tryptophan-rich domain [TRD] region in Table 2.

Discussion: The oat seed-specific ESTs generated by Beattie and co-workers [Ref. 19] proved to be a valuable resource that allowed for the identification of "full-length" ESTs for the relevant Umn sequences [Table 2] and for the identification of a novel OT, OT154 [Table 2 and Figure 1].

The known distribution of loci for Umn287, Umn360 and Umn856 markers [Table 3], particularly in the KxO map, is consistent with the occurrence of at least 2 gene assemblages, putative oat *Ha*-loci. One is on LG 22_44+18 and the other on LG 29_43.

Mapping results for the CAPS markers of the OT154s encoded by GO582063 and GO581838 [Table 3] support the proposed oat *Ha*-loci. The GO582063 marker maps to the proposed locus on KxO LG 22_44+18, expanding this locus to four characterized sequences/genes. The GO582063 locus is placed with Umn360 and Umn856 on TxM LG 15 thereby generating a putative oat *Ha*-locus with 3 characterized sequences/genes. In addition, the GO581838 locus is associated with the "unlinked" Umn360x locus in TxM map generating a second putative oat *Ha*-locus with 2 characterized sequences in this map.

It is not possible to determine absolute sequence/gene order within these assemblages as oat mapping populations do not have sufficient numbers of recombinant inbred lines to allow for this level of fine mapping. Given that the *Ha*-type genes have appeared to have undergone independent duplications and/or deletions at different stages of the evolution of the cereals and grasses in general [e.g. Refs. 12 and 22], it is possible that the oat assemblages may differ in the number and type of genes that are present. Alternatively, current differences may simply be an artefact arising from a lack of appropriate markers. These issues might be addressed by genomic cloning of large segments of DNA using artificial chromosomes. This approach has been used with success in other cereals [e.g. Refs. 12 and 22].

Mapping results for the OT3B3T[GO582128] CAPS marker are less informative. This marker was only useful in the DxE mapping population as the KxO and TxM parents were monomorphic. In addition, it mapped in DxE with 4 GBS markers, 2 of which map to KxO LG 15 [N.A. Tinker, personal communication]. It is not clear if this OT3B3T marker represents a potential third *Ha*-locus or if this marker represents an independent locus. In wheat, additional genes encoding *Pinb*-like proteins are found on the homeologous chromosomes 7A, 7B, and 7D [e.g. Ref. 23]

The locus for Umn162, Umn249, and Umn753 on KxO LG 6 appears to be independent of the assemblages discussed above. In our view, these 3 markers are synonymous, do not show any association with other *Ha*-locus markers, and, in general, they show limited sequence identity with known *Ha*-locus sequences. They most likely represent non-allelic, highly divergent homologues of the *Ha*-locus sequences.

While there is substantial evidence for the role of Pina and Pinb in the determination of grain hardness/texture in wheat and for their antimicrobial properties [e.g. Refs. 1-9], the *in vivo* activities of the oat tryptophanins/vromindolines remain to be established. Given the differences in seed composition between wheat and oat [e.g. oil content], the biological basis of grain hardness is likely to be more complicated in oat. However, the oat tryptophanins and other proteins encoded by the putative *Ha*-loci have TRDs [Table 2] that are generally comparable in composition to those of the Pins and Gsp1. It is also interesting to note that, in the MN841801-1xNoble-2 map [Ref. 24], loci for Umn162, Umn249, Umn360 and Umn856 fall within the support intervals for two minor QTLs [on LGs MN6 and MN14] for partial resistance to crown rust pathogen *Puccinia coronata*.

Acknowledgements: The authors would like to acknowledge superb technical support from Doris Luckert and the assistance of Dr. Nick Tinker with the mapping of loci and for access to data prior to their publication. Funding to M.A.T. from the Research Policy Committee of Cape Breton University, and to S.J.M. from Agriculture and Agri-Food Canada is recognized and greatly appreciated.

References: [1] C.F. Morris (2002), Pl. Mol. Biol. 48: 633-647, [2] M. Bhave and C.F. Morris (2008), Pl. Mol. Biol. 66: 205-219, [3] M. Bhave and C.F. Morris (2008), Pl. Mol. Biol. 66: 221-231, [4] L. Dubreil *et al*. (1998), Pl. Sci. 138: 121-135, [5] W. Jing *et al*. (2003), J. Bacteriology 185: 4938-4947, [6] R. Capparelli *et al*. (2005), Pl. Mol. Biol. 58: 857-867, [7] M. Faize *et al*. (2004), Pl. Sci. 167: 347-354, [8] L. Luo *et al*. (2008), Mol. Biol. Rep. 35: 195-200, [9] R.L. Phillips *et al*. 2011, J. Cer. Sci. 53: 112-117, [10] M.-F. Gautier *et al*. (1994), Pl. Mol. Biol. 25: 43-57, [11] M.-F. Gautier *et al*. (2000), Pl. Sci. 153: 81-91, [12] N. Chantret *et al*. (2005), The Pl. Cell 48: 633-647, [13] S. Rouvès *et al.* (1996), Pl. Breeding 115: 511-513, [14] B. Beecher *et al*. (2001), Theor. Appl. Genet. 102: 833-840, [15] M.C. Simeone and D. Lafiandra (2005), J. Cer. Sci. 41:115-122, [16] M.A. Tanchak *et al*. (1998), Pl. Sci. 137: 173-184, [17] S. Altschul *et al*. (1990), J. Mol. Biol. 215: 403-410, [18] B.T. Hizbai *et al*. (2012), The Pl. Genome 5: 164-175, [19] A.D. Beattie *et al*. (2008), Poster presentation, 8th International Oat Conference, Minneapolis, Mn, [20] N.A. Tinker *et al*. (2009), BMC Genomics 10: 39, [21] D.L. De Koeyer *et al*. (2004), Theor. Appl. Genet. 108: 1285-1298, [22] M. Charles et al. (2009), Mol. Biol. Evol. 26: 1651-1661, [23] F. Chen et al. (2010), Theor. Appl. Genet. 120: 745-751, and [24] V.A. Portyanko et al. (2005), Theor. Appl. Genet. 111: 313-324.