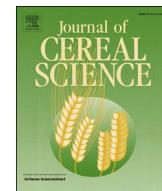




Contents lists available at SciVerse ScienceDirect

## Journal of Cereal Science

journal homepage: [www.elsevier.com/locate/jcs](http://www.elsevier.com/locate/jcs)

## Avenin diversity analysis of the genus *Avena* (oat). Relevance for people with celiac disease

D.M. Londono <sup>a,\*</sup>, W.P.C. van't Westende <sup>a,b</sup>, S. Goryunova <sup>c</sup>, E.M.J. Salentijn <sup>a,b</sup>, H.C. van den Broeck <sup>b</sup>, I.M. van der Meer <sup>b,d</sup>, R.G.F. Visser <sup>a</sup>, L.J.W.J. Gilissen <sup>b,d</sup>, M.J.M. Smulders <sup>a,b,d,\*</sup>

<sup>a</sup> Wageningen UR Plant Breeding, P.O. Box 386, NL-6700 AJ Wageningen, The Netherlands

<sup>b</sup> Plant Research International, P.O. Box 16, NL-6700 AA Wageningen UR, Wageningen, The Netherlands

<sup>c</sup> Vavilov Institute of General Genetics, Gubkin str. 3, 119991, Moscow GSP-1, Russia

<sup>d</sup> Allergy Consortium Wageningen, P.O. Box 16, NL-6700 AA Wageningen, The Netherlands

## ARTICLE INFO

## Article history:

Received 17 January 2013

Received in revised form

25 March 2013

Accepted 28 March 2013

## Keywords:

Gluten-free

Celiac disease

R5 antibody

G12 antibody

## ABSTRACT

Oat is widely consumed by people with celiac disease (CD). Its safety has been disputed because two peptides from oat avenins can be recognized as T cell epitopes by some CD patients. Differential signals of gluten-specific monoclonal antibodies and in-vitro T cells to oat varieties have suggested the existence of differences in immunogenicity. We aimed to clarify the nature of such responses by cloning avenin genes from 13 *Avena* species. A single oat plant contained up to 10 avenin genes. Avenin proteins clustered in four groups of which two contained the two avenin CD epitopes. All *Avena* species examined harbored avenins of these two groups, and as a consequence all contained avenins with the two avenin-specific epitopes, which makes it very unlikely to find oat cultivars that are devoid of these sequences. The established gluten epitopes from wheat, rye and barley were not present in oat avenins; some variants with two and three amino acid substitutions occurred, but they were predicted not to resist proteolysis in the gastro-intestinal tract. Perfect recognition sites of antibodies R5 and G12 were also not present in avenins. Thus, their signals to oat should not be interpreted as differences in immunogenicity for CD patients.

© 2013 Elsevier Ltd. All rights reserved.

### 1. Introduction

Celiac disease (CD) is an intolerance to gluten proteins (prolamins) from wheat, rye and barley that leads to a chronic inflammation of the small intestine and affects about 1% of Western population. The only remedy for CD patients is a gluten-free diet. There is an increasing demand for gluten-free products to diversify this diet and oat is considered an interesting alternative because it contains healthy compounds (essential amino acids, unsaturated fatty acids beta-glucans, polyphenols) that can supplement the diet of patients and help to prevent diabetes and vascular diseases (Andon and Anderson, 2008).

It has generally been accepted that CD patients can consume oats without detrimental inflammation of the small intestine (Thompson, 2003). A systematic study with 52 patients revealed that most of them, whether in remission or newly diagnosed, can add moderate amounts (about 50 g/day) of oats to their gluten-free diet without any harmful effects (Janatuinen et al., 1995). Another study described how children with CD tolerated oats in their gluten-free diet for a period of two years in which they were regularly monitored (Koskinen et al., 2009). Pulido et al. (2009) analyzed the studies published between 1995 and 2009, and concluded that the majority of adults and children with CD can tolerate moderate amounts (from 20 g/day for children to 70 g/day for adults) of pure oats. Probably the largest problem with oats for CD patients is the fact that most commercial oat products are contaminated with wheat (Gelinas et al., 2008). Contamination can start in the field when wheat or barley plants grow among oat plants and are not removed systematically, but contamination can also occur later on in the food production chain, during storage or processing. Consequently, a guaranteed gluten-free oat production chain is an essential requirement for CD-safe oat production. Such a chain exist in several Nordic countries (Sweden, Finland) where

\* Corresponding authors. Wageningen UR Plant Breeding, P.O. Box 16, NL-6700 AA Wageningen, The Netherlands. Tel.: +31 317 480840.

E-mail addresses: Diana.londono@wur.nl (D.M. Londono), wendy.vantwestende@wur.nl (W.P.C. van't Westende), orang2@yandex.ru (S. Goryunova), elma.salentijn@wur.nl (E.M.J. Salentijn), hetty.busink@wur.nl (H.C. van den Broeck), Ingrid.vandermeer@wur.nl (I.M. van der Meer), richard.visser@wur.nl (R.G.F. Visser), luud.gilissen@wur.nl (L.J.W.J. Gilissen), rene.smulders@wur.nl (M.J.M. Smulders).

70% of patients eat oats regularly as a healthy component of their gluten-free diet (Salovaara et al., 2010) and it was recently established in The Netherlands. Oat products containing less than 20 ppm gluten are allowed to be labeled and sold as “gluten-free” in Europe, because of its general safety and its contribution to a healthy diet.

Medical research has produced an extensive list of epitopes intrinsically present in wheat, rye and barley that are recognized by T cells from CD patients (Sollid et al., 2012). In wheat, HLA-DQ2/8 T cell epitopes occur in various gluten protein families:  $\alpha$ - and  $\gamma$ -gliadins, LMW- and HMW-glutenins. Presence and number of T cell epitopes per gluten molecule differ among members of these multigene storage protein families (van Herpen et al., 2006) and among individual cultivars (Spaenij-Dekking et al., 2005), due to naturally occurring amino acid substitutions in the epitopes (Mitea et al., 2010). Oat contains only one family of prolamins called avenins, which makes up 10–15% of the total seed protein content, in contrast to wheat, in which prolamins represent 80% of the total seed protein content. Four cases of oat avenin-sensitive CD patients have been described and the response has been attributed to three specific avenin-derived peptides that are different from all known T cell epitopes from wheat, rye and barley (Arentz-Hansen et al., 2004; Lundin et al., 2003; Vader et al., 2003). Recently a new nomenclature for CD epitopes was released based on unified criteria that the peptides should accomplish in order to be considered CD epitopes (Sollid et al., 2012). This list only includes two avenin-specific epitopes present in oats that can induce T cells.

Recently, a new discussion emerged based on responses of monoclonal antibodies (mAbs) and T cell clones to oat extracts *in vitro*. Silano et al. (2007) found that avenins from four oat varieties activated peripheral T cell clones from ten CD children on a gluten-containing diet. Comino et al. (2011) found that protein extracts from two of three groups of commercial oat varieties induced G12 mAb binding, and especially variety OM719 was able to induce proliferation and IFN-gamma production of peripheral blood mononuclear T cells from patients who were on a gluten-containing diet. The reactions were not due to contamination of the oats used. Mujico et al. (2011) observed differences in the binding of mAbs raised against a wheat LMW-glutenin derived peptide to oat cultivars. A possible relation was recently suggested between oat consumption and the persistence of duodenal intraepithelial lymphocytosis in CD patients on a gluten-free diet (Tuire et al., 2012).

These results raised the question whether toxicity in oats may have been overlooked, which would have implications on whether promotion of oat consumption by CD patients is justified.

**Table 1**

Summary of avenin genes obtained from genomic DNA of diploid, tetraploid, and hexaploid *Avena* species.

<i>Avena</i> species	Ploidy	Sample <sup>a</sup>	Cat#	Genome <sup>b</sup>	# of avenin genes	# of pseudogenes	# of genes with avenin epitopes
<i>A. canariensis</i>	2x	121	k292	Ac	6	0	3
<i>A. clauda</i>	2x	72	CN 19217	Cp	6	1	3
<i>A. ventricosa</i>	2x	53	CN21992	Cv	5	1	3
<i>A. longiglumis</i>	2x	534	k87	Al	5	0	4
<i>A. strigosa</i>	2x	127	k5196	As	7	0	6
<i>A. pilosa</i>	2x	148i	CN73755	Cp	5	0	2
<i>A. prostrata</i>	2x	86	2055	Ap	5	0	2
<i>A. damascena</i>	2x	57	CN 19457	Ad	6	0	2
<i>A. macrostachya</i>	4x	142	—	CmCm	4	1	3
<i>A. murphy</i>	4x	542	2088	AC <sup>2</sup>	8	1	4
<i>A. magna</i>	4x	123	k1852	AC <sup>2</sup>	6	0	2
<i>A. insularis</i>	4x	138	—	AC <sup>2</sup>	5	1	3
<i>A. sativa</i> cv Gigant	6x	—	—	ACD	10	0	4

<sup>a</sup> ID of the gDNA of diploid and tetraploid *Avena* species in the Vavilov Institute of General Genetics, Moscow, Russia.

<sup>b</sup> Genome AC(CD) of tetraploid species is represented by AC in this article.

Alternative explanations of these results include cross-reactivity of mAbs with other peptides that are not necessarily immunogenic for CD patients and promiscuity of T cells that have been induced by wheat gluten and that cross-react with variants of the canonical T cell epitopes (Anderson et al., 2000; Vader et al., 2003). Unfortunately, only few avenin sequences are known (Chesnut et al., 1989), and diversity within a single oat variety has not been studied, so it is unknown whether peptides in avenins could cause such responses.

The objective of this study was to determine the frequency of avenin-specific T cell epitopes, and whether T cell epitopes known from wheat, rye or barley are present in avenins. We also aimed to determine variation in copy number and sequence similarity among genes, and to estimate diversity between varieties and species. To achieve these objectives, we cloned and sequenced avenin genes present in the commercial hexaploid oat cultivar ‘Gigant’ and in accessions of 12 diploid and tetraploid *Avena* species, representing various A and C genomes within the genus *Avena* (Loskutov, 2008). We also analyzed an EST library from another hexaploid oat cultivar, ‘Dancer’.

## 2. Materials and methods

### 2.1. Plant material

Genomic DNA of twelve diploid and tetraploid *Avena* species representing the major genomes in *Avena* was obtained from the Vavilov Institute of General Genetics, Moscow, Russia (Table 1). We also included the hexaploid *Avena sativa* cultivar ‘Gigant’ from which genomic DNA (gDNA) was extracted using the DNeasy Plant Mini kit (Qiagen, Venlo, The Netherlands).

Grains in two different developmental stages (watery and early milk stage) were collected from the hexaploid *A. sativa* ( $2n = 6 \times = 42$ ) cv. ‘Gigant’ in an agricultural production field located in Munnekezijk, Groningen, The Netherlands. The hulls were removed to classify the grains according to developmental stage. The naked grains were frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  until processing. Messenger RNA was extracted using TRIzol and the RNeasy Plant Mini kit (Qiagen, Venlo, The Netherlands), and cDNA was synthesized using the iScript cDNA Synthesis Kit (Bio-Rad, Veenendaal, The Netherlands).

### 2.2. Primer design

The genes for storage proteins in cereals are often present in multiple copies per genome, which may be highly similar to each

other. To obtain avenins from oat, we implemented a strategy used to reliably obtain gamma-gliadin sequences from wheat despite the high similarity between various gene family members, only accepting a sequence if it was obtained in two independent PCR reactions (Salentijn et al., 2012). Primers were designed on conserved regions at the 5' and 3' ends of the coding sequence of 5 avenin genes described by Chesnut et al. (1989) and on avenin ESTs from 'Dancer' available in EMBL/GenBank, which assembled into 11 contigs of >99% identity (Table 2). Forward primers were designed on a part of the avenin signal peptide MKTFLI. Reverse primers were designed on the 3' end of the avenin genes, the stop codon TAA being present in the center or at the 3' end of the reverse primer. The primers were synthesised by Biolegio, Nijmegen, The Netherlands. In total, 7 different combinations of the following forward (F) and reverse (R) primers were used:

F2 5'-ATGAAGAACCTCCTCATC-3'  
 F4 5'-ATGAGGACCTCCTCATC-3'  
 Fd1 5'-ATGAARACMTTCYCCTCATC-3'  
 Fd2 5'-ATGAAGAMCTTYCTCATC-3'  
 R1 5'-AGTGTTCTTAAAGGCCAC-3'  
 R2 5'-AGTCITCTCAGAAGGCCAC-3'  
 R3 5'-AATCTTCTTAGCAACCAC-3'  
 R6 5'-CCCTTATAGTCTTCTCAG-3'  
 Rd2 5'-CYCTTRTAGTSTTCTYAG-3'

### 2.3. Cloning and sequencing of avenin genes from gDNA and cDNA

The set of seven primer pairs was used to amplify avenins from cDNA of cv. Gigant and gDNA of diploid and tetraploid *Avena* species (Table 1). Each reaction contained 50 ng gDNA or 2 µl cDNA, 75 mM Tris–HCl pH 9.0, 20 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.01% (w/v) Tween-20, 2.5 mM MgCl<sub>2</sub>, 250 µM of each dNTP, 200 nM of specific primer or 1200 nM of degenerate primer, and 1 U Goldstar Taq (Eurogentec, Liege, Belgium). The PCR reactions were performed in a Dyad Thermal Cycler PTC220 (Bio-Rad) starting with an initial denaturation step for 5 min at 94 °C followed by 1 cycle at 94 °C for 1 min, 56 °C for 1 min and 72 °C for 2 min. The 56 °C annealing temperature was subsequently reduced by 1 °C for the next 9 cycles, and continued at 47 °C for the remaining 26 cycles. A final extension followed at 72 °C for 10 min.

The PCR products, containing a mixture of gene sequences, were ligated into the pGEM®-T Easy Vector (Promega Benelux, Leiden, The Netherlands) and used for transformation of *E. coli* XL1-blue supercompetenT cells (Stratagene Europe, Amsterdam, The Netherlands). The white colonies were picked and grown overnight at 37 °C in LB-freezing medium into 96-well flat bottom plates. Each colony PCR contained 1 µl of overnight culture, 75 mM Tris–HCl pH 9.0, 20 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.01% (w/v) Tween-20, 1.5 mM MgCl<sub>2</sub>, 100 µM of each dNTP, 250 nM of both M13 forward and reverse primer and 0.5 U Goldstar Taq (Eurogentec). PCR was performed in a PTC200 (MJ Research, Bio-Rad) with an initial denaturation step at 94 °C for

5 min followed by 35 cycles of 94 °C for 30 s, 55 °C for 30 s and 72 °C for 2 min and a final extension step of 72 °C for 10 min. The colony PCR products were sequenced with M13 forward and reverse primers using the DYEnamic ET Terminator Cycle Sequencing Kit (Amersham Biosciences Benelux, Roosendaal, The Netherlands) on an ABI3710 automated sequencer (Applied Biosystems). On average, 16 clones per product were sequenced.

### 2.4. Analysis of sequences

The traces were checked manually and subsequently assembled using Lasergene 8.0 (DNAStar, Madison, WI). The mismatches within each contig were checked manually on the forward and reverse traces. Only sequences containing SNPs that were supported by two or more clones from at least two different PCR events were considered as genuine. Sequences obtained from cDNA and gDNA of cv. Gigant were assembled into 7 and 10 contigs of >99% identical sequences, respectively. Sequences obtained from diploid and tetraploid *Avena* species assembled into 41 contigs of >99% identical sequences; some contigs contained identical avenin genes from different *Avena* species. The sequences reported in this paper have been deposited in EMBL under Accession numbers FN706441–FN706450, HE576937–HE576976, and HE801178–HE801213.

An EST library containing 231 sequences of cv. 'Dancer' found in NCBI on May 2009 was analyzed also. The sequences with accession numbers [NCBI: G0584576.1, G0583621.1 and G0584748.1] were removed from the data set because they did not have the expected size and did not share the common features with the other sequences, e.g. the characteristic presence of eight cysteine residues in the translated amino acid sequence. All sequences were translated into amino acid sequences and used to perform a Bootstrap Neighbor-Joining test (1000 replicates) using MEGA 4.

### 2.5. Screening for CD toxic epitopes

All translated avenin protein sequences were screened for the presence of celiac-immunogenic or -toxic peptides (epitopes) described up to date in the literature (Solidi et al., 2012) as well as all possible variants of these epitopes in oats with 1, 2 or 3 amino acid substitutions. In silico digestion of avenin proteins was done using PeptideCutter (<http://expasy.org/>) taking into account the specificities of trypsin, pepsin (pH 1.3, and pH > 2), and chymotrypsin (high and low specificity).

### 2.6. SDS-PAGE

Oat avenins were extracted from 100 mg of commercial oat flour by addition of 0.5 ml of 50% (v/v) aqueous iso-propanol with continuous mixing (MS1 Minishaker, IKA Works, Inc.) at 1000 rpm for 30 min at room temperature, followed by centrifugation at 10,000 rpm for 10 min at room temperature. The residue was re-

**Table 2**

T cell epitope variants present in oat avenins that are predicted to resist trypsin and chymotrypsin digestion. T cell epitopes from wheat, rye and barley were not present in oat avenins, but variants with one, two, and three amino acid differences (in red and underlined) were present in avenin sequences. Of the 89 proteins of Fig. 3, only the eight variants listed here resisted in silico trypsin and chymotrypsin proteolysis. None of them remained completely intact if pepsin was also added.

T cell epitopes Deamidated	Epitope variants present in oat avenins that are predicted to be resistant to trypsin and chymotrypsin digestion							
DQ2.5-glia2 IQPEQPAQL (from wheat)	N I	Q Q	P P	Q Q	Q Q	Q L	A P	Q Q
DQ2.5-glut-L2 FS <u>QQQQ</u> ESPF (from wheat)	V	<u>Q</u>	Q	Q	Q	Q	P	F Y
DQ2.5-hor-1 PFPQPEQPF (from barley)	P	<u>Y</u>	P	E	<u>Q</u>	Q	P	F
DQ2.5-sec-1 PFPQPEQPF (from rye)	P	<u>Y</u>	P	E	<u>Q</u>	Q	P	F
DQ2.5-ave-1a PYPEQ <u>E</u> EPF (from oat)	P	<u>Y</u>	P	E	<u>Q</u>	Q	P	I
DQ2.5-ave-1b PYPEQEQPF (from oat)	P	<u>Y</u>	P	E	<u>Q</u>	Q	S	I
	P	<u>Y</u>	P	E	<u>Q</u>	Q	Q	L

extracted once with 0.5 ml 50% (v/v) aqueous iso-propanol. The two obtained supernatants were combined and can be considered to be the complete avenin protein extract, as shown by van den Broeck et al. (2011). The avenin protein content was quantified using the Bio-rad Protein Assay (Bio-Rad), based on the Bradford dye-binding procedure, according to manufacturer's instruction with BSA as a standard. Avenins were separated on SDS-PAGE gels (11%) using a Hoefer SE 260 mighty small II system (GE Healthcare) followed by staining with PageBlue (Fermentas, Thermo Scientific, St. Leon-Rot, Germany).

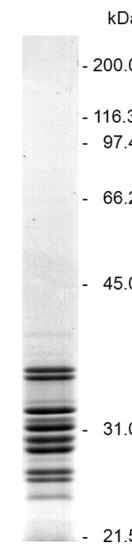
### 3. Results

#### 3.1. The avenins of oat

The general structure of avenin proteins is depicted schematically in Fig. 1. Similar to other prolamins, they have a high proline and glutamine content, low content of lysine, and they are insoluble in water. In total, 717 avenin sequences were obtained from gDNA of 13 diploid, tetraploid, and hexaploid *Avena* species containing the three major genomes (A, D, C) that occur within the genus. These sequences assembled in 83 contigs of >99% identical sequences, conservatively representing 78 different genes and 5 pseudogenes (with internal stop codons). Some contigs were composed of sequences from different species. The number of genes obtained from diploid and tetraploid accessions ranged from five to seven in diploid species, and from four to eight in tetraploid species. The number of genes obtained from gDNA of the hexaploid *A. sativa* cultivar 'Gigant' was 10 (Table 1).

Gene expression was studied in kernels of 'Gigant' in early stage of development. From cDNA we cloned and sequenced 120 PCR products (from two independent reactions) that assembled into only 7 contigs of >99% identical sequences (606–810 bp in length), indicating that 7 of the 10 different avenin genes obtained from gDNA were expressed during early development. It is likely that during later developmental stages of 'Gigant' a few more genes will be expressed since SDS-PAGE analysis of proteins from mature seeds of 'Gigant' showed 12 clear protein bands (Fig. 2). The EST library of cultivar 'Dancer' assembled into 11 contigs (S1).

Diploid species (*Avena vetricosa* and *Avena clauda*) and tetraploid species (*Avena macrostachya*, *Avena murphy* and *Avena insularis*) contained one sequence with internal stop codons (Table 1). According to an unrooted neighbor joining analysis, avenin genes clustered into four groups (I–IV; Fig. 3), with *Avena* species representative of all genomes in each group. From this we deduce that,

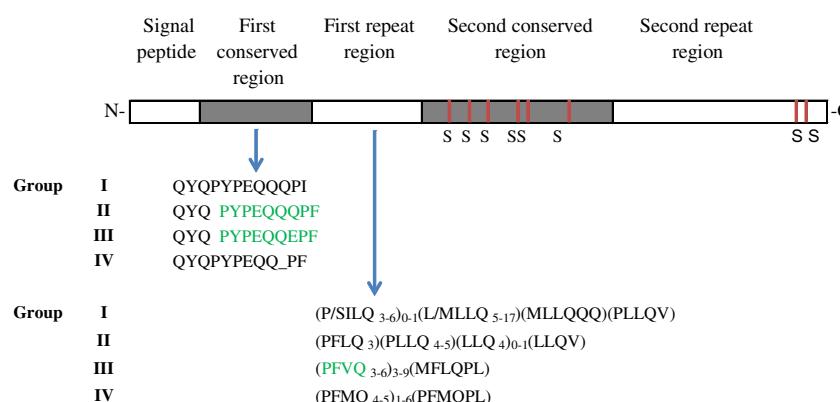


**Fig. 2.** SDS-PAGE (11%) analysis of an avenin extract from *A. sativa* cultivar Gigant followed by PageBlue staining. The apparent weight on the gel is somewhat higher than the molecular weight of avenins of 'Gigant' (21–31 kDa) as predicted based on the sequences.

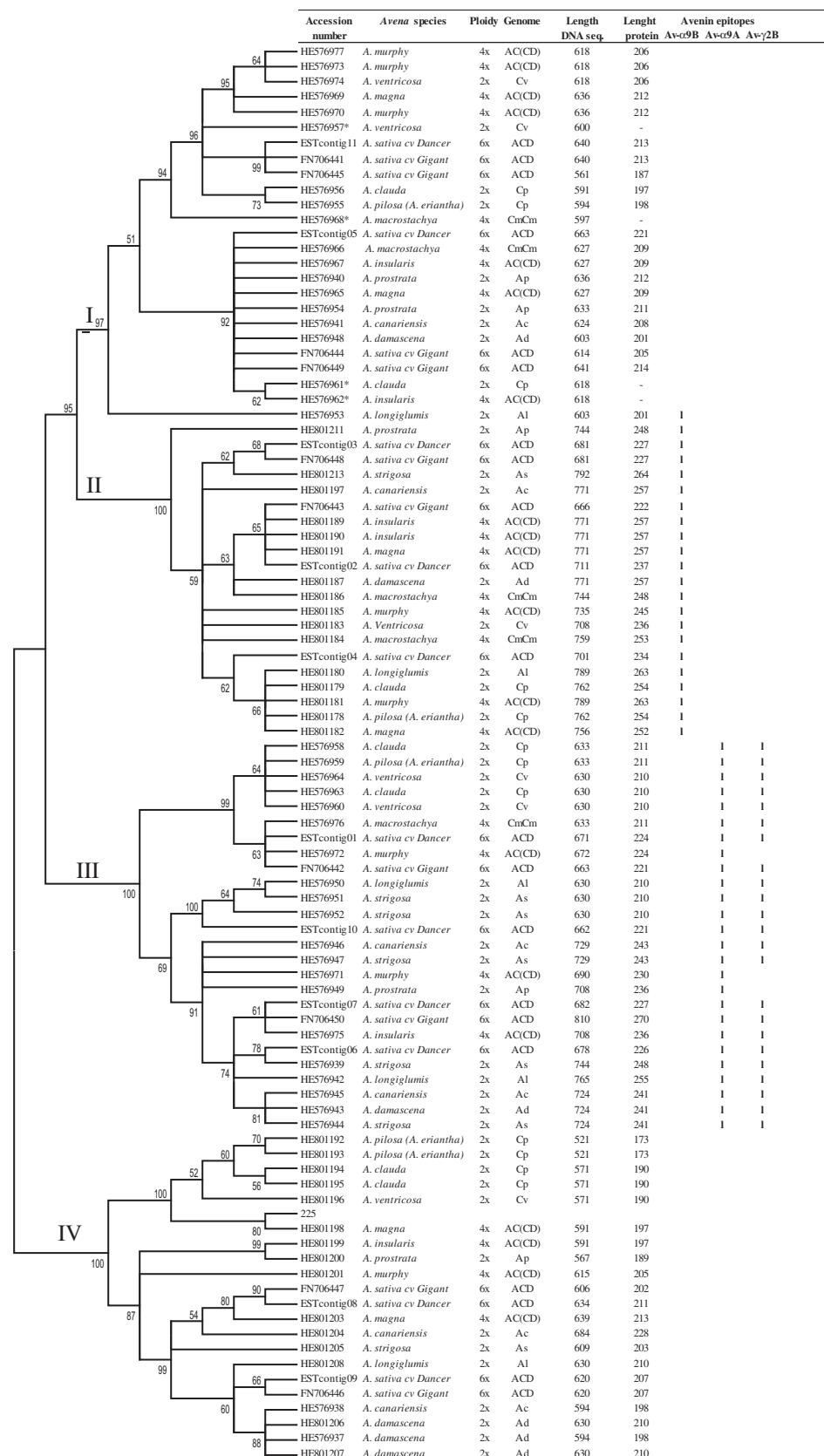
even if we did not have cloned all members of the gene family from each accession, we did cover the diversity present in the gene family across the genus. The N-terminal sequences exhibited a very high homology (S2), which is in accordance with the results presented by Pernollet et al. (1987). The length of the predicted proteins ranged from 187 to 270 AA. Differences in length were mainly due to presence of repetitive motifs within non-conserved regions (between positions 34–89 and 161–301 after the signal peptide). The predicted molecular weight of the avenins of 'Gigant' is 21–31 kDa, which is low in comparison to prolamins from wheat, rye and barley whose molecular weights range from 25 to 100 kDa (Tatham et al., 2000).

#### 3.2. Presence of celiac disease epitopes in avenins

To estimate the T-cell stimulatory capacity of oat avenins, all predicted avenin proteins were screened for the presence of known CD T cell epitopes [(Sollid et al., 2012) and variants thereof, as some substitutions may not abolish T cell activation (Mitea et al., 2010)]. Next, the potential protease resistance of the deduced peptides was



**Fig. 1.** Schematic structure of avenin proteins with typical amino acid motifs indicated. Gray areas correspond to highly conserved domains. Vertical red lines represent cysteines (S) conserved throughout avenin proteins. The location of the avenin epitopes according to current (Sollid et al., 2012) and old nomenclature (Vader et al., 2003) are indicated in green within the motifs. After Shewry et al. (1995). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 3.** Unrooted Neighbor Joining (NJ) tree of avenin genes from diploid, tetraploid and hexaploid *Avena* species. The avenin genes of *Avena* species with different ploidy levels formed four groups of sequences that are indicated as I–IV. Bootstrap values based on 1000 replicates are indicated (as %). Sequences containing avenin epitopes Av- $\alpha$ 9B and Av- $\alpha$ 9A (Solid et al., 2012) and Av- $\gamma$ 2B (Vader et al., 2003) are also indicated on the corresponding column for each epitope. The epitopes from wheat, rye and barley are absent in avenin sequences of the species analyzed.

**Table 3**

Variants of the recognition sites of R5 and G12 antibodies with one and two amino acid substitutions. No perfect recognition sites were present in oat avenins, but some variants with one or two amino acid substitutions (in red and underlined) were found.

Antibody	Reported antibody recognition sites	Variant in oat avenins with one (R5) or two (G12) amino acid substitution					% sequences containing the variant	Linked to avenin epitopes
G12	QQPFP	Q	Q	P	F	<u>L</u>	23.4	+
		Q	Q	P	F	<u>V</u>	27.6	+
	QQPY	Q	Q	P	F	<u>M</u>	23.4	—
		Y	Q	P	Y	<u>P</u>	100	—
	QPQLPY	Q	P	Q	L	<u>Q</u>	73.4	+
		Q	P	Q	Q	<u>Q</u>	40.4	+
	QPQQPY	Q	Q	Q	Q	<u>Q</u>	48.9	+
		Q	<u>Q</u>	Q	Q	<u>P</u>	14.9	—
	QPQLPF	Q	P	Q	Q	<u>L</u>	6.4	—
		Q	P	Q	L	<u>Q</u>	8.5	+

analyzed. Finally, the occurrence of recognition sites for the R5 and the G12 antibodies was verified.

The avenin-specific T cell epitopes PYPEQQQPF (Av- $\alpha$ 9B) and PYPEQQEPF (Av- $\alpha$ 9A) were present and occurred only once per protein. They were located in the first conserved region of avenin proteins: Av- $\alpha$ 9B was present in all genes of group II, whereas Av- $\alpha$ 9A occurred in all genes of group III. The motif QQPFPVQQQQPFPVQQ, which used to be considered as an epitope (Av- $\gamma$ 9B) but according to the new nomenclature it does not fulfill all criteria (Solid et al., 2012), was located immediately downstream of Av- $\alpha$ 9A in the first repetitive region in most of the sequences of group III (Fig. 1, S2). Avenins of groups I and IV did not contain the canonical avenin-specific T cell epitopes but variants thereof: PYPEQQPFM in proteins of group IV, and PYPEQQQPI or PYPEQQQSI in proteins of group I. As a consequence, Av- $\alpha$ 9B and Av- $\alpha$ 9A were found in avenins of the A/D and the C genomes, and in each of the 13 *Avena* species studied. Four of 10 avenin genes of hexaploid cultivar 'Gigant' contained at least one of the canonical avenin-specific T cell epitopes. In 'Dancer' 7 of 11 transcripts encoded one of these T cell epitopes (3 contained Av- $\alpha$ 9B and 4 contained Av- $\alpha$ 9A).

None of the gluten T cell epitopes from wheat, rye and barley were detected in avenins, but some variants with one, two and three amino acid substitutions were found (Table 2). The two variants with one amino acid substitution corresponded to the already described avenin T cell epitopes. Interestingly, the variants with two and three substitutions were predicted to be proteolysed in the gastrointestinal tract when we performed the analysis using the known specificities of the three enzymes, trypsin, chymotrypsin, and pepsin. In this analysis also, the two described avenin epitopes did not completely resist proteolysis as the phenylalanine was removed, leaving QPYPEQQQ and QPYPEQQEP peptides. If pepsin is omitted from the analysis, some variants are predicted to survive digestion (Table 2).

We further screened our set of avenin proteins for the 5-mer and 6-mer amino acid sequences that are recognized by the R5 antibody and the G12 antibody, respectively, which are used to detect gluten in an ELISA test. The R5 antibody recognizes QQPFP, LQPFP, LQPFP, QLPYP, and, to a lesser extent, QLPTF, QQSFP, QQTFP, PQPFP, QQPY, and PQPFP (Valdes et al., 2003), with an overall recognition motif of Q(P)-Q(P,L)-Q(L)-PF/YP (Osman et al., 2001). The G12 antibody recognizes the sequences QPQLPY, QPQQPY and QPQLPF (Moron et al., 2008). All those antibody recognition sequences were absent in avenin proteins. Variants of these recognition sites were present: with one amino acid difference for the R5 recognition site (Table 3), and with at least two amino acid differences for the G12 recognition site (Table 3). Two out of four R5-related variants and four out of six G12-related variants are present in sequences that also contain the avenin-specific epitopes, so it might be possible that they relate to the number of those epitopes but this relation

should be proved. All those variants are predicted to be degraded by pepsin, trypsin and chymotrypsin.

#### 4. Discussion

This work presents, for the first time, a thorough evaluation of the sequence diversity of avenin genes across the entire genus *Avena* (oat). We have studied the presence of celiac disease DQ2/8 T cell epitope sequences of wheat, barley and rye (according to current international agreement (Solid et al., 2012)), the occurrence of the two avenin-specific CD epitopes (Arentz-Hansen et al., 2004), and the existence of recognition sequences of several mAbs that are applied in commercially available gluten detection kits. The results enable to discuss deeply the safety of oats for patients suffering from celiac disease and the limitations of *in vitro* immunogenicity data in heterologous plant material.

##### 4.1. Avenin gene number

According to the estimations in oat made by Chesnut et al. (1989) based on Southern hybridization, and in line with large numbers of the homologous gliadin genes in wheat (van Herpen et al., 2006), at least 25 avenin genes (and pseudogenes) were expected to be present in oat. However, we detected only 10–11 avenin genes in the hexaploid cultivars 'Gigant' and 'Dancer'. It is possible that some cultivars contain loci with more genes through additional gene duplications. Alternatively, we may have underestimated the number of genes as we conservatively combined all sequences with 99% or higher similarity into one gene contig, thus possibly lumping highly similar genes. On the other hand, indirect evidence that the gene family is not very large comes from the fact that we did not find any pseudogenes in hexaploid *A. sativa* cultivars, and only very few in the sequences from other *Avena* species (overall 5 of 78 distinct avenin genes, or 6%). Pseudogenes may be the hallmark of an expanded gene family of storage proteins (in wheat, 87% of the  $\alpha$ -gliadins were pseudogenes (van Herpen et al., 2006)). It is not likely that our primers systematically would have missed particular groups because EST sequences clustered in the same groups.

A low number of avenin genes compared to gluten genes in wheat is consistent with the difference in the prolamin fraction of the storage protein composition (10–15% of the total seed protein content in oats, compared to about 80% in wheat (Tatham et al., 2000)). The low number of proteins is supported by the similar number of bands on an SDS-PAGE gel. van den Broeck et al. (2011) extracted avenins from a commercial oat flour and distinguished 12 protein bands (SDS-PAGE) and found that each band corresponded to a single avenin protein on a 2D-gel.

#### 4.2. Avenin gene family expansion

A neighbor-joining analysis showed that the avenin gene sequences of the entire *Avena* genus representing the diploid, tetraploid and hexaploid species and genomes (A/D and C) (Loskutov, 2008) obtained in this paper, clustered in four groups (Fig. 3). The genes in the four groups differed from each other by sequence differences in the two conserved regions, and by differences in repeat motifs and the number of repeat units in the two repetitive regions (S2). The codon-based test for selection (Z-test) showed evidence for purifying selection for genes of each of the four groups. Each group consists of avenins from the A and C genomes of the diploid species, and of tetraploid and hexaploid *Avena* species. This indicates that the most recent common ancestor (MRCA) of the *Avena* genus most likely contained one or more representatives of each of the four avenin groups before the A and C genomes diverged. In the Groups III and IV, two genome-specific subgroups appear to exist: one subgroup with sequences from genome A and the other one with sequences from genome C. Repeat motifs within the first and the second repetitive regions of group IV contained methionine while the others did not. Chesnut et al. (1989) regarded this group as a separate protein subfamily. Absence of genome-specific subgroups in group II could be due to a lack of initial divergence (in agreement with the low diversity within this group) and/or incomplete lineage sorting. Presence of two A-genome subgroups in group III, each of them containing sequences of *Avena strigosa* and *Avena longiglumis*, may indicate a specific duplication of sequences of this group in the MRCA of at least some of the A genome diploid species. The allohexaploid species *A. sativa* contains avenins from each of the four groups and subgroups except the C-genomic subgroup of group IV, indicating that the allohexaploid has maintained all groups brought in by the three genomes.

#### 4.3. Phylogenetic origin

There are some structural differences between avenins and other prolamins (especially those from wheat, barley and rye) regarding polymorphism, molecular weight and number and location of repetitive domains and cysteine residues (Shewry et al., 1995; Tatham et al., 2000). Alignment of the second conserved region of the avenins against EMBL/GenBank sequences (not shown) indicated that the highest degree of homology are gamma-type S-rich prolamins, such as the gamma-secalins in barley and the gamma-hordeins in rye that all contain eight cysteine residues at similar positions. Similarly to secalins, none of the avenin sequences contained tryptophan. We confirmed the presence of aspartic acid in position 7 and its replacement by asparagine as shown by Pernollet et al. (1987), or, in some avenins, by tyrosine.

Thus, genus-wide cloning and sequencing of the avenins clearly showed the limited range of diversity of the avenin gene family. With this approach, for the first time, a complete picture of the avenins in oats has been obtained, as a basis to establish the safety of oat for CD patients.

#### 4.4. The safety of oats for people with celiac disease

The safety and the diet-improving effect of oats for the majority of celiac patients is currently well-recognized (Pulido et al., 2009). It is fair to say that a large intervention study (although not intended as such) regarding the safety of oats is currently ongoing in the form of daily pragmatic practice of oat consumption in Scandinavia and The Netherlands, where Celiac patient societies stimulate oat-based food consumption from established gluten-free oat production chains, and where patients have a positive attitude on the gluten-free oat products. Seventy percent of CD

patients in Nordic countries and an increasing number of patients in The Netherlands eat oats regularly with full contentment (Salovaara et al., 2010). In addition, the advantages of oat consumption by diabetic individuals, as CD and diabetes regularly happen to coincide, are welcomed (personally communicated by the board of the Dutch Diabetic Society).

Toxicity problems of oat products, in general, are related to contamination with gluten, and, in some rare cases, to the presence of the two avenin-specific sequences harboring the epitopes Av- $\alpha$ 9A (PYPEQQEPF) and Av- $\alpha$ 9B (PYPEQQQPF) that can induce an immune response in a few celiac patients (Arentz-Hansen et al., 2004; Lundin et al., 2003; Vader et al., 2003). Although the number of oat-sensitive patients appears to be very low, this aspect attained much attention and definitely caused reluctance in oat consumption by the celiac population. Peptides containing these two avenin epitopes turned out to be common in oat (Fig. 3).

#### 4.5. Antibody and T cell tests

Based on our extensive set of deduced avenin proteins we were, for the first time, able to determine that the 17 internationally accepted CD epitopes from wheat, rye and barley that are recognized by CD4+ T cells (Sollid et al., 2012), as well as the p31–43 epitope that has been implicated in the innate immune response, are all absent in oat avenins. Variants of these epitopes do occur in oats, but they differ at two or three of the nine amino acid residues. None of them is predicted to survive digestion by pepsin, trypsin and chymotrypsin. Thus, they are unlikely to be clinically relevant. The two avenin-specific epitopes described by Arentz-Hansen et al. (2004), to which few CD patients specifically react, were found in two of the four groups of avenins that occur in each of the genomes. Hence, these epitopes are probably present in all *Avena* species. Based on in silico proteolysis we predict that they also represent the most abundant protease-stable peptides. So, the question is what could cause the differential reaction of monoclonal antibodies and peripheral T cells *in vitro* to oat varieties. None of the recognition sites of the R5 and the G12 antibody tests are present in avenins. Only variant sequences with one (in the case of R5) or two (in the case of G12) amino acid residue difference were detected, some of which were linked to the avenin-specific epitopes (Table 3). Thus, the immunogenic responses that are described to oat extracts of some varieties (Comino et al., 2011; Mujico et al., 2011) are most likely the result of cross-reactivity with sequences present in avenins with one (R5) or two (G12) amino acid difference and this should be considered to be beyond the identified clinically toxic sequence profile (Sollid et al., 2012). Hence, the relation between the antibody signal and a T cell response as established for gliadins in wheat and prolamins in rye and barley should not simply be extrapolated to oat.

#### 4.6. Limitations of test systems

Thus, test systems that do not directly detect the intact epitopes are limited in their applicability, as the correlation between the antibody signal and the presence of intact epitopes as established for gliadins in wheat and prolamins in rye and barley, should not be extrapolated beyond these species (Sollid et al., 2012). This would appear obvious, but antibodies developed to detect gluten in wheat were used on the proven gluten-safe crops like the dicot Quinoa (Zevallos et al., 2012), which is not even a cereal. Without confirmation by other methods, such data are misleading to the overall patient population. Here, the scientific world should take its responsibility, and should be reluctant in sending out signals that might cause commotion among the patient society before relevance and extent of possible harm are reliably demonstrated.

## 5. Conclusion

The genus-wide genomics analysis of the avenin gene family with regard to celiac disease (CD) has shown that none of the internationally agreed epitopes from gluten of wheat, barley and rye are present in oat avenins. The presence of the two described avenin-epitopes was confirmed within a conserved avenin region in all *Avena* species examined, making it unlikely to find oat varieties that lack these sequences. Therefore, if responses of T cells of CD patients to oat extracts occur, they are most likely toward these two peptides. The monoclonal antibody signals (R5 and G12) are probably due to cross-reactive or promiscuous recognition of avenin peptides, and unless the relation with the two avenin-epitopes is proven, such signals should not be interpreted as differences in immunogenicity of oat varieties for CD patients.

## Acknowledgments

This research was funded by the Celiac Disease Consortium, an Innovative Cluster approved by the Netherlands Genomics Initiative and partially funded by the Dutch Government (BSIK03009), by the Ministry of Economic Affairs (KB05-001-019, KB05-003-032, KB15-001-007), and The Europees Fonds voor Regionale Ontwikkeling (EFRO), which is supported by the European Union and the province Gelderland

## Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.jcs.2013.03.017>.

## References

- Anderson, R.P., Degano, P., Godkin, A.J., Jewell, D.P., Hill, A.V.S., 2000. In vivo antigen challenge in celiac disease identifies a single transglutaminase-modified peptide as the dominant A-gliadin T-cell epitope. *Nature Medicine* 6, 337–342.
- Andon, M., Anderson, J., 2008. The oatmeal-cholesterol connection: 10 years later. *American Journal of Lifestyle Medicine* 2, 51–57.
- Arentz-Hansen, H., Fleckenstein, B., Molberg, O., Scott, H., Koning, F., Jung, G., Roepstorff, P., Lundin, K.E.A., Sollid, L.M., 2004. The molecular basis for oat intolerance in patients with Celiac disease. *Plos Medicine* 1, 84–92.
- Chesnut, R.S., Shotwell, M.A., Boyer, S.K., Larkins, B.A., 1989. Analysis of avenin proteins and the expression of their messenger-RNAs in developing oat seeds. *Plant Cell* 1, 913–924.
- Comino, I., Real, A., de Lorenzo, L., Cornell, H., López-Casado, M.Á., Barro, F., Lorite, P., Torres, M.I., Cebolla, Á., Sousa, C., 2011. Diversity in oat potential immunogenicity: basis for the selection of oat varieties with no toxicity in coeliac disease. *Gut* 60, 915–922.
- Gelinas, P., McKinnon, C.M., Mena, M.C., Mendez, E., 2008. Gluten contamination of cereal foods in Canada. *International Journal of Food Science and Technology* 43, 1245–1252.
- Janatuinen, E.K., Pikkarainen, P.H., Kemppainen, T.A., Kosma, V.M., Jarvinen, R.M.K., Uusitupa, M.I.J., Julkunen, R.J.K., 1995. A comparison of diets with and without oats in adults with celiac disease. *New England Journal of Medicine* 333, 1033–1037.
- Koskinen, O., Villanen, M., Korponay-Szabo, I., Lindfors, K., Maki, M., Kaukinen, K., 2009. Oats do not induce systemic or mucosal autoantibody response in children with coeliac disease. *Journal of Pediatric Gastroenterology and Nutrition* 48, 559–565.
- Loskutoff, I.G., 2008. On evolutionary pathways of *Avena* species. *Genetic Resources and Crop Evolution* 55, 211–220.
- Lundin, K.E.A., Nilsen, E.M., Scott, H.G., Loberg, E.M., Gjoen, A., Bratlie, J., Skar, V., Mendez, E., Lovik, A., Kett, K., 2003. Oats induced villous atrophy in coeliac disease. *Gut* 52, 1649–1652.
- Mitea, C., Salentijn, E.M.J., van Veelen, P., Goryunova, S.V., van der Meer, I.M., van den Broeck, H.C., Mujico, J.R., Montserrat, V., Gilissen, L.J.W.J., Drijfhout, J.W., Dekking, L., Koning, F., Smulders, M.J.M., 2010. A universal approach to eliminate antigenic properties of alpha-gliadin peptides in celiac disease. *Plos One* 5, e15637.
- Moron, B., Cebolla, A., Manyani, H., Alvarez-Maqeda, M., Megias, M., Thomas, M.D.C., Lopez, M.C., Sousa, C., 2008. Sensitive detection of cereal fractions that are toxic to celiac disease patients by using monoclonal antibodies to a main immunogenic wheat peptide. *American Journal of Clinical Nutrition* 87, 405–414.
- Mujico, J.R., Mitea, C., Gilissen, L.J.W.J., de Ru, A., van Veelen, P., Smulders, M.J.M., Koning, F., 2011. Natural variation in avenin epitopes among oat varieties: implications for Celiac disease. *Journal of Cereal Science* 54, 8–12.
- Osman, A.A., Uhlig, H.H., Valdes, I., Amin, M., Mendez, E., Mothes, T., 2001. A monoclonal antibody that recognizes a potential toxic repetitive pentapeptide epitope in gliadins. *European Journal of Gastroenterology & Hepatology* 13, 1189–1193.
- Pernollet, J.-C., Huet, J.-C., Galle, A.-M., Sallantin, M., 1987. N-terminal sequences of oat avenins compared to other cereal prolamins. *Biochimie* 69, 683–689.
- Pulido, O., Gillespie, Z., Zarkadas, M., Dubois, S., Vavasour, E., Rashid, M., Switzer, C., Godefroy, B.S., 2009. Introduction of oats in the diet of individuals with celiac disease: a systematic review. *Advances in Food and Nutrition Research* 57, 235–285.
- Salentijn, E.M.J., Mitea, D.C., Goryunova, S.V., Van der Meer, I.M., Padoleau, I., Gillissen, L.J.W.J., Koning, F., Smulders, M.J.M., 2012. Celiac disease T cell epitopes from gamma-gliadins: immunoreactivity depends on the genome of origin, transcript frequency, and flanking protein variation. *BMC Genomics* 13, 277.
- Salovaara, H.O., Kanerva, P., Loponen, J., Sontag-Strohm, T.S., 2010. Oats in gluten-free diet, considerations of a pragmatic policy. In: Second International Symposium on Gluten-free Cereal Products and Beverages. University of Helsinki, Tampere, Finland, pp. 193–194.
- Shewry, P.R., Tatham, A.S., Barro, F., Barcelo, P., Lazzari, P., 1995. Biotechnology of breadmaking: unravelling and manipulating the multi-protein gluten complex. *Biotechnology* 13, 1185–1190.
- Silano, M., Di Benedetto, R., Maioratti, F., De Vincenzi, A., Calcaterra, R., Cornell, H.J., De Vincenzi, M., 2007. Avenins from different cultivars of oats elicit response by celiac peripheral lymphocytes. *Scandinavian Journal of Gastroenterology* 42, 1302–1305.
- Sollid, L.M., Qiao, S.W., Anderson, R.P., Gianfrani, C., Koning, F., 2012. Nomenclature and listing of celiac disease relevant gluten T-cell epitopes restricted by HLA-DQ molecules. *Immunogenetics* 64, 455–460.
- Spaenij-Dekking, L., Kooy-Winkelhaar, Y., Van Veelen, P., Drijfhout, J.W., Jonker, H., Van Soest, L., Smulders, M.J.M., Bosch, D., Gilissen, L., Koning, F., 2005. Natural variation in toxicity of wheat: potential for selection of nontoxic varieties for celiac disease patients. *Gastroenterology* 129, 797–806.
- Tatham, A.S., Gilbert, S.M., Fido, R.J., Shewry, P.R., 2000. Extraction, separation and purification of wheat gluten proteins and related proteins of barley, rye and oats. *Methods in Molecular Medicine* 41, 55–73.
- Thompson, T., 2003. Oats and the gluten-free diet. *Journal of the American Dietetic Association* 103, 376–379.
- Tuire, I., Marja-leena, L., Teea, S., Katri, H., Jukka, P., Päivi, S., Heini, H., Markku, M., Pekka, C., Katri, K., 2012. Persistent duodenal intraepithelial lymphocytosis despite a long-term strict gluten-free diet in celiac disease. *The American Journal of Gastroenterology* 107, 1563–1569.
- Vader, L.W., Stepien, D.T., Bunnik, E.M., Kooy, Y.M.C., De Haan, W., Drijfhout, J.W., Van Veelen, P.A., Koning, F., 2003. Characterization of cereal toxicity for celiac disease patients based on protein homology in grains. *Gastroenterology* 125, 1105–1113.
- Valdes, I., Garcia, E., Llorente, M., Mendez, E., 2003. Innovative approach to low-level gluten determination in foods using a novel sandwich enzyme-linked immunosorbent assay protocol. *European Journal of Gastroenterology & Hepatology* 15, 465–474.
- van den Broeck, H.C., Gilissen, L.J.W.J., Smulders, M.J.M., van der Meer, I.M., Hamer, R.J., 2011. Dough quality of bread wheat lacking alpha-gliadins with celiac disease epitopes and addition of celiac-safe avenins to improve dough quality. *Journal of Cereal Science* 53, 206–216.
- van Herpen, T., Goryunova, S.V., van der Schoot, J., Mitreva, M., Salentijn, E., Vorst, O., Schenk, M.F., van Veelen, P.A., Koning, F., van Soest, L.J.M., Vosman, B., Bosch, D., Hamer, R.J., Gilissen, L.J.W.J., Smulders, M.J.M., 2006. Alpha-gliadin genes from the A, B, and D genomes of wheat contain different sets of celiac disease epitopes. *BMC Genomics* 7, 1.
- Zevallos, V.F., Ellis, H.J., Suligoj, T., Herencia, L.I., Ciclitira, P.J., 2012. Variable activation of immune response by quinoa (*Chenopodium quinoa* Willd.) prolamins in celiac disease. *The American Journal of Clinical Nutrition* 96, 337–344.