Can an Increase in Celiac Disease Be Attributed to an Increase in the Gluten Content of Wheat as a Consequence of Wheat Breeding?

Donald D. Kasarda

Western Regional Research Center, Agricultural Research Service, U.S. Department of Agriculture, 800 Buchanan Street, Albany, California 94710, United States

ABSTRACT: In response to the suggestion that an increase in the incidence of celiac disease might be attributable to an increase in the gluten content of wheat resulting from wheat breeding, a survey of data from the 20th and 21st centuries for the United States was carried out. The results do not support the likelihood that wheat breeding has increased the protein content (proportional to gluten content) of wheat in the United States. Possible roles for changes in the per capita consumption of wheat flour and the use of vital gluten as a food additive are discussed.

KEYWORDS: gluten, celiac disease, breeding, wheat protein content

INTRODUCTION

There is recent evidence that the incidence of celiac disease has increased during the second half of the 20th century.1−3 These studies provide evidence that the incidence is changing based on the presence of tissue transglutaminase antibodies (a marker for celiac disease) in selected sets of serum samples, although there are no continuous data relating to the incidence of celiac disease in the U.S. population on a year-by-year basis. In addition, there is increasing interest in other gluten-related disorders: wheat allergy and nonceliac gluten sensitivity.4 It has been speculated that the increase in celiac disease may have occurred because of changes in wheat proteins that resulted from wheat breeding—mainly an increase in the gluten content,4 which is directly proportional to protein content. Here, I will focus on conventional breeding carried out for various purposes: to increase or decrease gluten proteins or modify them in other ways, to increase yield, to change kernel size or shape, or to improve disease or insect resistance. My focus will be on the United States, and because there are no GMO-type (genetically engineered) wheats used commercially in the United States, direct genetic modification of the wheat genome to increase protein content need not be considered. I will discuss briefly the history of wheat from its domestication to modern times and some factors that may have a bearing on the question of whether or not the increase in celiac disease can be attributed to an increase in the gluten content of modern wheats.

WHEAT HISTORY

Diploid Wheat. Man first domesticated a diploid and tetraploid wheats about 10000 years ago5 (see timeline of Figure 1). One of the wheat forms originally domesticated, sometimes called “einkorn,” had only one genome, usually designated the A genome. Because the vegetative tissues of plants contain two copies of the genome (hence, diploid), the plant designation is AA. The diploid wild wheat that was first domesticated is thought to have been Triticum urartu, but other diploid wheat species may have been involved as well. These diploid species are likely to have evolved from a common line that included oats, barley, and rye (Figure 2). In the centuries following domestication, the genome changed as a consequence of natural selection and hybridization with other species such as oat, barley, and rye.

Figure 1. Events in evolution in relation to the appearance of wheat (approximate dates).

Figure 2. Divergence of a common line leading to the diploid progenitors of wheat.
of breeding and selection, which would have introduced and
fixed some new genes and/or controlling sequences into the
genome. Consequently, the domesticated modern diploid
wheat took on a significantly different character from the wild
wheat. Wild wheat species usually have tiny grains, often
needle-like in appearance, compared to the larger grains of
modern wheats. When grown today, the grains of these wild
species usually have high protein contents, in the 16–28% range. Early farmers and breeders probably selected for seed
size (among other traits)—larger grains being easier to recover
from the plant during threshing. In doing so, they inadvertently
increased the starch content relative to protein content; thus,
the apparent seed protein percentage decreased. The
domesticated diploid wheat, (einkorn) was designated Triticum
monococcum (genus Triticum, species monococcum) by Linnaeus
in the 18th century. The small amount of T. monococcum that
is grown today is used mostly for animal feed. In the early
millennia following domestication, the protein content of
domesticated wheat was most likely steadily declining because
of selection for traits that had an inverse relationship to protein
percentage, such as seed size and starch content. Throughout most of the 10000 years of wheat domestication
it was not possible to measure the protein content of wheat
grain. The concept of protein as a unique substance was
developed only about 250 years ago, and it was not until 1883
when Johan Kjeldahl developed his method for organic
nitrogen determination that the stage was set for the
development of an accurate and moderately convenient method
for determining the protein content of wheat. Even so, the
proper factor for converting the nitrogen content of wheat grain
to protein content was a subject of debate throughout the early
part of the 20th century. In the late 20th century, with the
development of the near-infrared reflectance (NIR) spectroscopy
method for the determination of wheat protein content,
measurement of this quantity was greatly simplified. The gluten
content of wheat is approximately proportional to the protein
content and usually ranges between 70 and 75% of total protein
content. Statements in the literature, such as, “Since the first
farmers, strains of wheat have been selectively bred for their
gluten content, and the gluten content has progressively
increased through time” and “…the selection of wheat varieties
with higher gluten content has been a continuous process
during the past 10,000 years…” are therefore not correct. The
earliest farmers were selecting for many traits, but protein
content was not one of them. However, with the development of yeast-fermented (leavened) bread baking about 2000–5000
years ago (the date is not known with any precision), farmers
may have indirectly begun to select for protein content because
leavened bread requires a relatively high protein content (at
present 11% is considered somewhat minimal for bread making
in the United States).

**Polyploid Wheats: Tetraploids and Hexaploids.** It is
likely that tetraploid wheat was domesticated about the same
time as diploid wheat. The tetraploid wheats have two
genomes that, for vegetative tissues, are designated AABB
(hence, tetraploid) (Figure 3). Wheats with more than one
genome are known as polyploid wheats. (I will not attempt a
discussion of polyploid formation here.) The A genome of the
tetraploid wheats is closely similar to that of T. monococcum,
and the B genome is related to Triticum speltaeoides. When grown
today, wild tetraploids usually have protein contents in the
range of 16–27%, whereas domesticated tetraploids usually
have lower protein contents of about 10–12%. The wild
tetraploids are often designated Triticum dicoccoides, and the
domesticated equivalents are classified as Triticum turgidum
(sometimes Triticum dicoccum). T. turgidum has two sub-
species/varieties, one with the common name emmer and the
other called durum. Emmer wheat is not free-threshing, which
means that there is a tightly adhering husk, or glume, that is
difficult to remove from the grains. Durum wheat is free-
threshing, that is, the seeds are readily released from the
glumes. The main wheat of the Roman empire was emmer, and
its current form, now somewhat in vogue for gourmet cooking,
is usually known as farro. Durum wheats are now used mainly
for pasta, although bread can be made from durum wheat, and
such bread is common in parts of Italy, such as the Puglia
region.

The next step up in complexity was to hexaploid wheat,
which, for Triticum aestivum, consists of three genomes
designated A, B, and D. The A and B genomes of hexaploid
wheats are nearly identical to the A and B genomes of
tetraploid wheat. The hexaploid wheats have no wild
equivalents, and the ABD hexaploids resulted from hybrid-
ization of a cultivated (domesticated) emmer and a wild grass
species known as Triticum tauschii (sometimes called goat
grass) followed by polyploid formation to give rise to a new
species having three genomes designated AABBDD in
vegetative tissues (hence, hexaploid), as summarized in Figure
3 (heads of the diploid progenitor species are pictured in Figure
4). The ABD hexaploid wheats are free-threshing with the
exception of spelt wheat, which has tightly adhering glumes.
These hexaploids arose within a few thousand years of the first
wheat domestication.

**PROTEIN CONTENT AND USAGE OF WHEAT TYPES**

Most wheats used for breadmaking are ABD hexaploid wheats
with “hard” endosperm texture, but there are hexaploid wheat
varieties/cultivars with “soft” texture that are used mainly for
pastry and cake baking. Hard wheats have been selected for
higher protein content than soft wheats and, for breadmaking,
higher protein content (about 12–14%) is usually desirable.
For the soft wheats, low protein is usually desirable, because it
is the starch rather than the protein that plays the more
important role in determining desirable pastry characteristics.
Consequently, soft wheats have mostly been selected for low
protein content, which usually falls in the range of 7–11%. All-

---

**Figure 3.** Combinations of diploid wheats leading to the polyploid forms.
perspective flour has an intermediate range of protein, which makes it acceptable, but not optimal, for most uses.

There are few pertinent papers that address the question of whether or not the protein content of the U.S. wheat crop has increased over time in the 20th century. Although adequate data may exist, much of the literature from the first half of the 20th century seems not to have been digitized. Information on the Kansas crops of 1949–2011 shows that for the main wheat type grown in Kansas (hard winter wheat, used primarily for breadmaking), many crops fell within a protein content range of 11–13%, with an unusual high of 14.1% in 1956 and an unusual low of 10.7 in 1961. Some papers from the first half of the 20th century indicated protein contents ranging from 8.77 to 14.26% for various samples of hard winter wheats. The 8.77% number appeared to be an outlier, and there was no obvious deviation from the results for protein content in the second half of the 20th century.

The hard spring wheats, grown mostly in the Northern Plains, are considered to be highly desirable for bread baking and tend to have protein contents that in general exceed the usual protein contents of winter wheats by about 2 percentage points. The protein contents of hard spring wheats vary on average over a range of about 3–4 percentage points from year-to-year, usually falling in the 12–16% range as exemplified by the data in Table 1 (extracted from Table 42 of Bailey). The basis for the higher protein content of the spring wheat from the Northern Plains is probably related partly to genetics and partly to growing conditions (environment). In the late 1920s, the protein contents of the Northern Plains wheats were relatively low (about 13%), presumably due to normal to high levels of rainfall, but increased to average levels of about 15% in the 1930s, presumably due to drought conditions (Table 1). The North Dakota Wheat Commission reported that, in 2009, the hard red spring wheat crop...yielded an average of 13.1% (which) was well below the traditional level of more than 14%, and these protein contents are fairly typical of late 20th century crops for the hard spring wheat region. Various studies have compared the protein contents of wheat varieties from the early part of the 20th century with those of recent varieties. When grown under comparable conditions, there was no difference in the protein contents. Although nitrogen fertilization can have strong effects on protein content for some wheat varieties, the data do not seem to be in accord with the likelihood that recent fertilization protocols have had a strong effect on the protein contents of wheat grown in the United States.

Interpretation of protein data is complicated by occasional major deviations from the more usual range. In 1938, the protein content of spring wheat was exceptionally high (Table 1), averaging close to 19%; these years of exceptionally high protein (or low protein) occur occasionally and are likely to result mainly from environmental factors, rather than nitrogen fertilization or wheat breeding. To maintain a uniformity of quality characteristics from year to year, flour mills usually blend wheat flour that is intended for commercial use by specific customers, for example, bakeries. Very high protein content would usually be unsuitable for direct use, and so high-protein wheat flour would usually be blended with lower protein grain to achieve a more normal protein level before reaching the consumer. The connection between celiac disease and wheat ingestion was not made until 1950, so that any variations in the incidence of celiac disease prior to that would not have been distinguished from gastrointestinal diseases in general. Even today, data for the incidence of celiac disease that could be used to recognize short-term variations are not available.

With acknowledgment that the data that are available were not suitable for rigorous statistical interpretation, I found no evidence of any obvious trend toward higher protein content for either winter or spring wheats since the early part of the 20th century when the key ancestral varieties of current bread wheats (of both winter and spring habits) were introduced to the United States from Europe and Asia. Hard winter and spring wheats are the predominant types grown in the United States, mostly for bread baking, and these wheats will have the highest protein contents. Soft wheats have significantly lower

<table>
<thead>
<tr>
<th>crop year</th>
<th>no. of samples</th>
<th>av protein (%)</th>
<th>standard deviation (σ)</th>
<th>av moisture content (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1925</td>
<td>33246</td>
<td>12.49</td>
<td>1.34</td>
<td></td>
</tr>
<tr>
<td>1926</td>
<td>26145</td>
<td>13.28</td>
<td>1.55</td>
<td>13.7</td>
</tr>
<tr>
<td>1927</td>
<td>63944</td>
<td>11.96</td>
<td>0.78</td>
<td>13.2</td>
</tr>
<tr>
<td>1928</td>
<td>49964</td>
<td>12.42</td>
<td>0.77</td>
<td>13.4</td>
</tr>
<tr>
<td>1929</td>
<td>37202</td>
<td>13.70</td>
<td>1.41</td>
<td>13.4</td>
</tr>
<tr>
<td>1930</td>
<td>52041</td>
<td>14.85</td>
<td>1.47</td>
<td>13.1</td>
</tr>
<tr>
<td>1931</td>
<td>17182</td>
<td>15.00</td>
<td>1.22</td>
<td></td>
</tr>
<tr>
<td>1932</td>
<td>45027</td>
<td>14.21</td>
<td>0.99</td>
<td>11.7</td>
</tr>
<tr>
<td>1933</td>
<td>28829</td>
<td>15.03</td>
<td>0.89</td>
<td>11.5</td>
</tr>
<tr>
<td>1934</td>
<td>12900</td>
<td>14.80</td>
<td>1.04</td>
<td>11.4</td>
</tr>
<tr>
<td>1935</td>
<td>28544</td>
<td>15.30</td>
<td>1.71</td>
<td>11.8</td>
</tr>
<tr>
<td>1936</td>
<td>16698</td>
<td>15.92</td>
<td>1.64</td>
<td></td>
</tr>
<tr>
<td>1937</td>
<td>12185</td>
<td>14.83</td>
<td>1.28</td>
<td>11.6</td>
</tr>
<tr>
<td>1938</td>
<td>13169</td>
<td>18.78</td>
<td>1.04</td>
<td>11.5</td>
</tr>
</tbody>
</table>
protein contents because high protein contents are undesirable for many of the products made from soft wheats.

**DIET, GLUTEN CONSUMPTION, AND CELIAC DISEASE**

Bread is the most common wheat-based food in the United States, although pastries and pasta are also widely consumed. Because bread wheats in the United States range in protein content from about 11% to about 16% and pastry wheats from about 7% to 11%, the amounts of gluten ingested might vary considerably from individual to individual depending on diet choices. Economic Research Service (USDA-ARS-ERS) statistics indicate that the per person per year intake of wheat flour (from all classes of wheat) reached a high of 220 lb (100 kg) per person in about 1900, declining steadily to a low of about 110 lb (50 kg) per person in 1970 and then gradually rising to about 146 lb (66 kg) per person in 2000, with a slight decrease occurring since 2000 to about 134 lb (61 kg) per person in 2008 (Figure 5). The per capita numbers of Figure 5 do not include U.S. wheat that is exported or used for animal feed, but presumably include wheat flour used to produce vital gluten. If I arbitrarily suggest that the average protein content of the wheat crop (all types of wheat) would be around 11%, the gluten equivalent of this U.S. grain/flour would be about 11.1 lb (5 kg) per person (assuming a per capita intake of 134 lb of wheat flour and that 75% of total flour protein would be gluten protein); if the average protein content was closer to 12%, the gluten equivalent would be about 12.1 lb (5.1 kg) per capita.

**VITAL GLUTEN**

Gluten fractionated from wheat flour by washing starch granules from a dough (sometimes called vital gluten) is often added to food products to achieve improved product characteristics. About 80% of the gluten used in the United States is imported—mainly from Australia, the European Union, Canada, and China. The question of how much vital gluten contributes to the total consumption of gluten (wheat flour and wheat grain + vital gluten) is complicated by a lack of accessible information about gluten production and imports for the United States in recent years and by indications that imports are rising rapidly. I make a crude estimate of what both factors might be as follows: Gluten imports were $177 \times 10^6$ lb in 1997 and $386 \times 10^6$ lb (175 $\times 10^6$ kg) in 2007. A two-point linear extrapolation to 2012 indicates that imports would currently be $490 \times 10^6$ lb ($222 \times 10^6$ kg). I assumed that the population of the United States is $330 \times 10^6$, that vital gluten is 75% protein, and that 80% of vital gluten is used for human food, to obtain a per person (per year) intake of gluten of 0.9 lb (408 g). Following the same approach, I estimate the gluten intake in 1977 to be 0.3 lb (136 g) per person. Thus, it appears that vital gluten consumption has tripled since 1977. This increase is of interest because it is in the time frame that fits with the predictions of an increase in celiac disease. It is difficult to say whether or not this increase in vital gluten consumption might contribute to an increase in the incidence of celiac disease—particularly when compared to the much larger intake of gluten from the consumption of wheat flour (11–12 lb/5.0–5.5 kg) per person. Similarly, I note that, although wheat flour consumption seems to be decreasing slightly in recent years (Figure 5), there was an increase in the yearly consumption of wheat flour of about 35 lb (15.9 kg) per person in the period from 1970 to 2000, which would correspond to an additional 2.9 lb (1.3 kg) of gluten per person from that extra flour intake, so that the 1970 intake of 9.1 pounds (4.1 kg) of gluten (from flour or grain) had increased to 12 pounds (5.4 kg) in 2000. It may be noted that whole wheat products, which are increasing in consumption for health reasons (especially the higher fiber content), often have vital gluten added to them to compensate for the negative effects of the ground whole grain on quality factors, such as loaf volume in breadmaking. This increase amounts to about 1.5–2.0 percentage points in product protein content, but the significance (if any) of this increase is not known—it must be considered in the context of other factors.

There is evidence that the D genome of bread wheat has more epitopes active in celiac disease than the A and B genomes, and some of these might be the most active epitopes. Consequently, tetraploid wheats (mainly durum wheats used for pasta) and diploid wheats, such as T. monococcum, are likely to be less toxic to celiac patients than bread wheats. However, given that all of these wheats, even T. monococcum, have proteins with several of the potentially active sequences that have been defined as being toxic for people with celiac disease, the significance of the decreased number of epitopes in tetraploids and diploids needs additional investigation.

The estimates that celiac disease is increasing in the United States are based on studies that cover approximately the last half of the 20th century. Could this increase be attributed to the increased consumption of wheat during that period, or to an increased use of vital gluten in food products, rather than any systematic increase in the protein content of wheat in the United States? Some diagnosed celiac patients in remission exhibit changes to the intestinal epithelium characteristic of celiac disease with a daily intake of 50–100 mg of gluten, generally considered the minimal toxic dose of gluten. The average slice of bread weighs approximately 40 g and contains about 2.4 g of protein—of that protein, about 1.8 g (1800 mg) would be gluten. It might seem, intuitively, that the variations in protein content of wheat would not be a key factor in the sensitization of potential celiac patients, given the seemingly large excess of gluten in most wheat-containing products over the minimum allowable intake. Nevertheless, one cannot rule out that the process by which the immune system switches from tolerance of wheat gluten protein to intolerance (celiac disease) has been facilitated by increased gluten consumption.
disease) might be dependent on the total amount of gluten encountered. Development of immune system tolerance to gluten proteins in general is an as yet imperfectly understood process—as is the loss of immune system tolerance to gluten proteins that is characteristic of celiac disease.

In summary, I have not found clear evidence of an increase in the gluten content of wheat in the United States during the 20th century, and if there has indeed been an increase in celiac disease during the latter half of the century, wheat breeding for higher gluten content does not seem to be the basis. Changes in the per capita intake of wheat and gluten might play a role; both increased during the period in question, but there is a lack of suitable data on the incidence of celiac disease by year to test those possibilities. The normal fluctuation of wheat crop protein content from year-to-year, being equivalent to or larger than the intake of fractionated gluten, is a complicating factor, but one that would be diminished by the practice of flour blending. Other factors, such as per capita vital gluten intake, variations in individual diets with regard to the amount and types of wheat consumed, wheat genetics, and agronomic practices (such as nitrogen fertilization), that affect protein content might contribute to determining the “toxicity” of wheat for people with the appropriate genetic susceptibility for celiac disease (mainly those carrying the genes for particular proteins of the major histocompatibility complex, DQ2 and DQ8); further research would be needed to evaluate such factors.

**AUTHOR INFORMATION**

**Notes**

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

I thank the following for helpful comments on the manuscript: Susan B. Altenbach, Peter H. R. Green, Brett Carver, Frances M. Dupont, Gary Vocke, and Peter R. Shewry.

**REFERENCES**


