Analysis of Gluten Containing Species As Contaminants Through DNA Barcoding in Rice

Cassidy Chin[1] , Anaya Awais[2], Avni Sahni[1], Richard Li[3]

[1]New Explorations into Science Technology and Math High School, [2]Chapin School [3]John Jay College of Criminal Justice

URBAN BARCODE RESEARCH PROGRAM
Supported by the Pinkerton Foundation

Cold Spring Harbor Laboratory
DNA LEARNING CENTER

Abstract: This study investigates gluten-containing species that are present in rice products commonly assumed to be gluten-free to evaluate potential food safety concerns through DNA barcoding. The objective is to extract and analyze plant DNA from four rice types—as rice is naturally gluten free—jasmine (Thailand), brown (USA), whole grain brown (USA), and basmati (India)—to detect possible contamination with gluten-producing species such as wheat, barley, or rye. Cross-contamination or mislabeling can pose serious risks to individuals with celiac disease or gluten sensitivity. DNA barcoding was performed by extracting genomic DNA, preparing it through native ligation using Oxford Nanopore's sequencing protocol, and analyzing the amplified rbcL locus sequences. Overall, the results provide preliminary evidence that these rice samples are free of gluten, but further testing across a wider range of products is necessary before making conclusions.

Introduction:

Gluten is a group of proteins, which are found in the endosperm of wheat, barley, and rye[1]. In individuals with celiac disease, the immune system mistakenly recognizes gluten as a harmful substance[2], triggering an immune response that damages the lining of the small intestine[3-8]. The only effective treatment is strict adherence to a gluten-free diet. Rice, a staple grain consumed globally, is naturally gluten-free and is often recommended as a safe alternative for individuals with celiac disease. However, research has shown that even inherently gluten-free grains like rice can become contaminated with gluten during farming, milling, or packaging processes that are shared with gluten-containing grains. Currently, the United States Food and Drug Administration requires that gluten-free products contain less than 20 parts per million(ppm) of gluten, aiming to protect consumers[9]. However, studies have revealed that up to 40% of labeled gluten-free products may contain gluten above detectable thresholds, highlighting the need for more accurate and accessible testing methods[10-12]. This study specifically investigates the potential for gluten cross-contamination in rice samples, using DNA barcoding to detect gluten-containing plants. By focusing on different rice types from a variety of global sources, the goal is to determine whether international rice products maintain gluten-free integrity or show signs of contamination. The hypothesis is that despite being naturally gluten-free, some rice products may contain detectable gluten-containing plants due to cross-contact, posing risks for people who rely on strict gluten-free diets for medical reasons. DNA barcoding, a molecular technique, enables the detection of trace amounts of DNA by amplifying specific sequences using polymerase chain reaction (PCR). By applying DNA barcoding, this study aims to assess the prevalence of gluten cross-contamination in products from rice samples. While PCR-based DNA barcoding is highly effective, this study employed the nanopore sequencing (also known as NanoSequencing), a newer technology for DNA analysis. Nanopore sequencing works by passing single strands of DNA through tiny protein nanopores and measuring changes in electrical current to read the nucleotide sequence in real time. Unlike Sanger sequencing, nanopore technology is highly sensitive, allowing for the identification of gluten-containing species such as wheat, barley, or rye in mixed food samples.

Materials & Methods:

Rice sample were soaked in sterile water and allowed to sit at room temperature for several hours in order to allow gluten-containing particles that would remain on the rice's surface to transfer to the water it was being soaked in. After soaking, the mixture was strained to remove the solid rice grains, leaving behind a cloudy water solution containing suspended residue. This solution was then transferred into multiple centrifuge tubes and left undisturbed, allowing the particulate matter to settle at the bottom. The top layer of clear liquid was carefully removed, and the bottom layer, containing the residue, was used for DNA extraction. Plant DNA was extracted using the Chelex method. The amplification of the barcoding regions of the plant DNA was carried out using the primers of the rbcL locus according to the UBRP protocol (2024). The sequencing reagent used for library preparation was the Oxford Nanopore Native Barcoding Kit 24 (NBD-24), which was employed for DNA repair, barcode ligation, and adapter ligation (Figure 1). The sequencing was carried out using a MinION Flow Cell device (Oxford Nanopore Technologies, Figure 2). To identify the species, the sequences obtained were compared to the sequences in the GenBank nucleotide database using BLAST search.

Results & Discussion:

DNA barcoding results for Rices 1, 2, 3, and 6 indicated that all sequences were identified as species within the Oryza genus, with high sequence identity percentages ranging from 97.99% to 100.00%. No sequences matched with wheat (Triticum), barley (Hordeum), or rye (Secale), which are the main gluten-containing grains. These findings suggest the absence of gluten contamination across all four samples. Samples 4 and 5 did not produce sufficient sequence reads and were excluded from further analysis. Most queries matched to Oryza sativa, the domesticated Asian rice species, and Oryza rufipogon, its wild ancestor. In several samples, sequences also aligned with Oryza longistaminata, a wild African rice species. One notable outlier was observed in Rice 6, Query 13, which matched Leersia perrieri with 99.00% identity. Leersia perrieri is a wild grass species closely related to rice but not classified within the Oryza genus. Across all samples, query alignment confidence was consistently high, with nearly all matches exceeding 98% identity (Table 1-4). These data confirm the genetic identification of the samples as rice species and show no indication of gluten-containing plant DNA. The DNA barcoding results indicate that none of the tested rice samples—jasmine rice (Thailand), brown rice (USA), whole grain brown rice (USA), and basmati rice (India)—contained detectable DNA from gluten-containing grains such as wheat, barley, or rye. These findings support the conclusion that there was no gluten cross-contamination in the samples analyzed. This is especially relevant for individuals with celiac disease or gluten sensitivities, as even trace contamination can lead to adverse health effects. Our results also suggest that rice grown and processed in different parts of the world can remain free of gluten contamination. Rice 6 (Basmati rice) contains one query sequence that matched Leersia perrieri (see Table 4), a wild grass species that is closely related to rice but not part of the Oryza genus. While Leersia perrieri is not known to contain gluten, its presence could reflect minor environmental or field-level contamination, hybridization events, or technical artifacts during sequencing. Importantly, this result does not indicate the presence of gluten but does raise questions about the genetic purity or biodiversity within some rice products. It is important to note that this study was limited in scope, analyzing only four rice samples from specific brands and countries. The absence of gluten-containing plants in these particular samples does not guarantee that all rice products on the market are free from cross-contamination. Factors such as shared processing equipment, packaging environments, and transportation methods can vary significantly across brands and facilities, potentially introducing gluten in ways not detected here. Broader sampling across more rice types, brands, and regions would be needed to make a definitive conclusion about the overall gluten safety of rice products.

Figures & Tables

Figure 1. Flow chart of the process used to complete the procedure for the experiment.

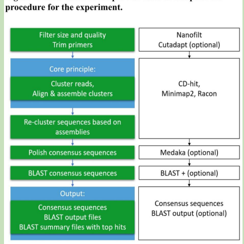

Figure 2. A MinION Mk1B or Mk1C device from Oxford Nanopore Technologies

Table 1: Results for the Species Identified in Jasmine Rice (Thailand)

Query ID	Percent Identification	Species Scientific Name	Contamination
1	99.83%	Oryza rufipogon	No
2	99.67%	Oryza sativa Indica Group	No
3	98.99%	Oryza rufipogon	No
4	99.33%	Oryza sativa Indica Group	No
5	99.17%	Oryza sative	No
6	99.50%	Oryza sativa Indica Group	No
7	99.67%	Oryza sativa Indica Group	No
8	99.83%	Oryza sativa Japonica Group	No
9	99.66%	Oryza sativa Indica Group	No
10	99.50%	Oryza sativa Indica Group	No
11	99.50%	Oryza rufipogon	No
12	99.67%	Oryza rufipogon	No
13	99.67%	Oryza rufipogon	No
14	99.33%	Oryza sativa Indica Group	No
15	98.67%	Oryza sativa Indica Group	No
16	99.66%	Oryza sativa Indica Group	No
17	99.83%	Oryza sativa Indica Group	No
18	99.17%	Oryza longistamina	No
19	99.57%	Oryza rufipogon	No
20	99.50%	Oryza sativa Indica Group	No
21	99.67%	Oryza sativa Indica Group	No
22	99.50%	Oryza sativa Indica Group	No
23	99.83%	Oryza rufipogon	No
24	99.67%	Oryza sativa Indica Group	No

Table 2: Results for the Species Identified in Brown Rice (USA)

Query ID	Percent Identification	Species Scientific Name	Contamination
1	99.33%	Oryza sativa Indica Group	No
2	99.83%%	Oryza rufipogon	No
3	99.67%	Oryza sativa Indica Group	No
4	99.67%	Oryza rufipogon	No
5	98.47%	Oryza rufipogon	No
6	99.17%	Oryza rufipogon	No
7	99.17%	Oryza sativa Indica Group	No
8	99.17%	Oryza rufipogon	No
9	99.83%	Oryza rufipogon	No
10	99.82%	Oryza sativa Indica Group	No
11	99.67%	Oryza rufipogon	No
12	99.83%	Oryza rufipogon	No
13	99.83%	Oryza rufipogon	No
14	99.67%	Oryza rufipogon	No
15	99.50%	Oryza longistaminata	No
16	98.50%	Oryza longistaminata	No
17	99.67%	Oryza sativa Indica Group	No
18	99.67%	Oryza sativa Indica Group	No
19	99.67%	Oryza rufipogon	No
20	99.67%	Oryza rufipogon	No
21	99.83%%	Oryza rufipogon	No
22	99.67%	Oryza sativa Indica Group	No
23	99.83%%	Oryza rufipogon	No
24	99.83%%	Oryza longistaminata	No

Table 3: Results for the Species Identified in Whole Grain Brown Rice (USA)

Query ID	Percent Identification	Species Scientific Name	Contamination
1	99.83%	Oryza rufipogon	No
2	99.50%	Oryza sativa Indica Group	No
3	99.67%	Oryza rufipogon	No
4	99.33%	Oryza sativa	No
5	99.33%	Oryza longistaminata	No
6	99.17%	Oryza sativa Indica Group	No
7	99.67%	Oryza rufipogon	No
8	99.83%	Oryza rufipogon	No
9	99.50%	Oryza rufipogon	No
10	99.67%	Oryza rufipogon	No
11	99.83%	Oryza rufipogon	No
12	99.67%	Oryza rufipogon	No
13	99.67%	Oryza rufipogon	No
14	99.83%	Oryza sativa Indica Group	No
15	99.67%	Oryza sativa Indica Group	No
16	99.33%	Oryza sativa Indica Group	No
17	100.00%	Oryza sativa	No
18	100.00%	Oryza sativa	No
19	99.67%	Oryza rufipogon	No
20	99.67%	Oryza rufipogon	No
21	99.50%	Oryza rufipogon	No
22	99.50%	Oryza rufipogon	No
23	98.66%	Oryza longistaminata	No
24	99.83%	Oryza rufipogon	No
25	99.83%	Oryza rufipogon	No
26	99.66%	Oryza rufipogon	No
27	99.83%	Oryza rufipogon	No
28	99.17%	Oryza sativa Indica Group	No
29	98.67%	Oryza rufipogon	No
30	99.17%	Oryza sativa Indica Group	No
31	99.17%	Oryza rufipogon	No
32	99.57%	Oryza sativa	No
33	99.83%	Oryza rufipogon	No
34	99.83%	Oryza rufipogon	No
35	99.17%	Oryza rufipogon	No
36	99.83%	Oryza rufipogon	No
37	98.67%	Oryza rufipogon	No
38	98.34%	Oryza rufipogon	No
39	99.83%	Oryza rufipogon	No
40	99.83%	Oryza rufipogon	No
41	99.83%	Oryza rufipogon	No

Table 4: Results for the Species Identified in Basmati Rice (India)

Query ID	Percent Identification	Species Scientific Name	Contamination
1	99.50%	Oryza rufipogon	No
2	100.00%	Oryza sativa	No
3	98.65%	Oryza rufipogon	No
4	99.83%	Oryza rufipogon	No
5	99.83%	Oryza rufipogon	No
6	99.50%	Oryza longistaminata	No
7	100.00%	Oryza sativa	No
8	100.00%	Oryza sativa	No
9	100.00%	Oryza sativa	No
10	99.66%	Oryza sativa	No
11	100.00%	Oryza sativa	No
12	99.66%	Oryza sativa	No
13	99.00%	Leersia perrieri	No
14	98.67%	Oryza rufipogon	No
15	99.84%	Oryza rufipogon	No
16	99.50%	Oryza rufipogon	No
17	100.00%	Oryza sativa	No
18	99.83%	Oryza sativa	No
19	100.00%	Oryza sativa	No
20	99.83%	Oryza rufipogon	No
21	99.83%	Oryza rufipogon	No
22	98.67%	Oryza rufipogon	No
23	99.83%	Oryza rufipogon	No
24	99.00%	Oryza rufipogon	No
25	99.83%	Oryza sativa	No
26	99.83%	Oryza rufipogon	No
27	99.83%	Oryza rufipogon	No
28	98.67%	Oryza rufipogon	No
29	99.44%	Oryza sativa	No
30	99.83%	Oryza rufipogon	No
31	99.84%	Oryza rufipogon	No
32	99.57%	Oryza sativa	No
33	100.00%	Oryza sativa	No
34	99.82%	Oryza sativa	No

References:

1.) Glutenin. Glutenin - an overview | ScienceDirect Topics. (n.d.).
2.) What is gluten?. Celiac Disease Foundation. (n.d.).
3.) Posner, E. B. (2023, August 8). Celiac disease. StatPearls [Internet].
4.) What is gluten and what does it do?. Johns Hopkins Medicine. (2024, June 20).
5.) Celiac disease: Symptoms & how it's treated. Cleveland Clinic. (2024a, May 3).
6.) Pediatric celiac disease - conditions and treatments. Children's National Hospital. (n.d.).
7.) Celiac disease patients with ongoing intestine damage at lymphoma risk. Columbia University Irving Medical Center. (2018, September 22).
8.) Celiac disease patients with ongoing intestine damage at lymphoma risk. Columbia University Irving Medical Center. (2018, September 22).
9.) Program, H. F. (n.d.). Questions and answers on the gluten-free food labeling final rule. U.S.
Food and Drug Administration.
10.) Wieser, H., Segura, V., Ruiz-Carnicer, Á., Sousa, C., & Comino, I. (2021, June 29). Food
Safety and cross-contamination of gluten-free products: A narrative review. Nutrients.
11.) Výrostková, J., Regecová, I., Zigo, F., Marcinčák, S., Kožárová, I., Kováčová, M., &
Bertová, D. (2022, July 7). Detection of gluten in gluten-free foods of plant origin. Foods (Basel,
Switzerland).
12.) Google. (n.d.). Genome sequencing using native ligation barcoding. Google Docs.

Acknowledgements: We would like to sincerely thank Dr. Feitzinger for her generous assistance with the sequencing process. Her expertise and contributions played a critical role in the success of this study. We would also like to thank Dr. Allison Mayle for her feedback and encouragement in this research.