






OPEN ACCESS

ORIGINAL ARTICLE

Metagenomics of the faecal virome indicate a cumulative effect of enterovirus and gluten amount on the risk of coeliac disease autoimmunity in genetically at risk children: the TEDDY study

Katri Lindfors ¹, Jake Lin,^{1,2} Hye-Seung Lee,³ Heikki Hyöty,¹ Matti Nykter,¹ Kalle Kurppa ^{1,4,5}, Edwin Liu,^{6,7} Sibylle Koletzko,^{8,9} Marian Rewers,¹⁰ William Hagopian,¹¹ Jorma Toppari,^{12,13} Annette-Gabriele Ziegler,^{14,15,16} Beena Akolkar,¹⁷ Jeffrey P Krischer,³ Joseph F Petrosino,¹⁸ Richard E Lloyd,¹⁸ Daniel Agardh ¹⁹, the TEDDY Study Group

For numbered affiliations see end of article.

Correspondence to

Katri Lindfors, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland; katri.lindfors@tuni.fi

KL and JL contributed equally.

Received 6 September 2019

Revised 1 November 2019

Accepted 5 November 2019

ABSTRACT

Objective Higher gluten intake, frequent gastrointestinal infections and adenovirus, enterovirus, rotavirus and reovirus have been proposed as environmental triggers for coeliac disease. However, it is not known whether an interaction exists between the ingested gluten amount and viral exposures in the development of coeliac disease. This study investigated whether distinct viral exposures alone or together with gluten increase the risk of coeliac disease autoimmunity (CDA) in genetically predisposed children.

Design The Environmental Determinants of Diabetes in the Young study prospectively followed children carrying the HLA risk haplotypes DQ2 and/or DQ8 and constructed a nested case–control design. From this design, 83 CDA case–control pairs were identified. Median age of CDA was 31 months. Stool samples collected monthly up to the age of 2 years were analysed for virome composition by Illumina next-generation sequencing followed by comprehensive computational virus profiling.

Results The cumulative number of stool enteroviral exposures between 1 and 2 years of age was associated with an increased risk for CDA. In addition, there was a significant interaction between cumulative stool enteroviral exposures and gluten consumption. The risk conferred by stool enteroviruses was increased in cases reporting higher gluten intake.

Conclusions Frequent exposure to enterovirus between 1 and 2 years of age was associated with increased risk of CDA. The increased risk conferred by the interaction between enteroviruses and higher gluten intake indicate a cumulative effect of these factors in the development of CDA.

INTRODUCTION

The incidence of autoimmune diseases is rising more rapidly than can be explained by genetics, supporting the role of environmental factors in the disease pathogenesis.¹ Coeliac disease, a dietary gluten-driven chronic small bowel enteropathy, is characterised by

Significance of this study

What is already known on this subject?

- ▶ Intake of high amounts of gluten increases the risk of coeliac disease in genetically predisposed children.
- ▶ Gastrointestinal infections have been associated with an increased risk of coeliac disease.

What are the new findings?

- ▶ The prospective metagenomics screening of the stool virome shows that the cumulative number of stool enteroviral exposures between 1 and 2 years of age are associated with an increased risk for coeliac disease autoimmunity.
- ▶ There is an interaction between cumulative enteroviral exposures between 1 and 2 years of age with cumulative gluten intake by 2 years of age in relation to the risk of coeliac disease autoimmunity.
- ▶ The effect of enteroviruses on the risk of coeliac disease autoimmunity is higher when greater amounts of gluten are consumed.

How might it impact on clinical practice in the foreseeable future?

- ▶ This study suggests that enteroviral infections early in life are potentiated by high gluten intake, which may trigger the development of coeliac disease autoimmunity in genetically predisposed children.
- ▶ More studies are needed to evaluate potential pathogenetic mechanisms of this interaction, which could offer new opportunities for the development of preventive strategies for coeliac disease.

an autoimmune response against tissue transglutaminase (tTG). The main autoantigen in coeliac disease is tTG, which post-translationally deamidates gluten-derived gliadin peptides.² Coeliac disease autoimmunity (CDA), which refers to the appearance of



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Lindfors K, Lin J, Lee H-S, *et al.* Gut Epub ahead of print: [please include Day Month Year]. doi:10.1136/gutjnl-2019-319809

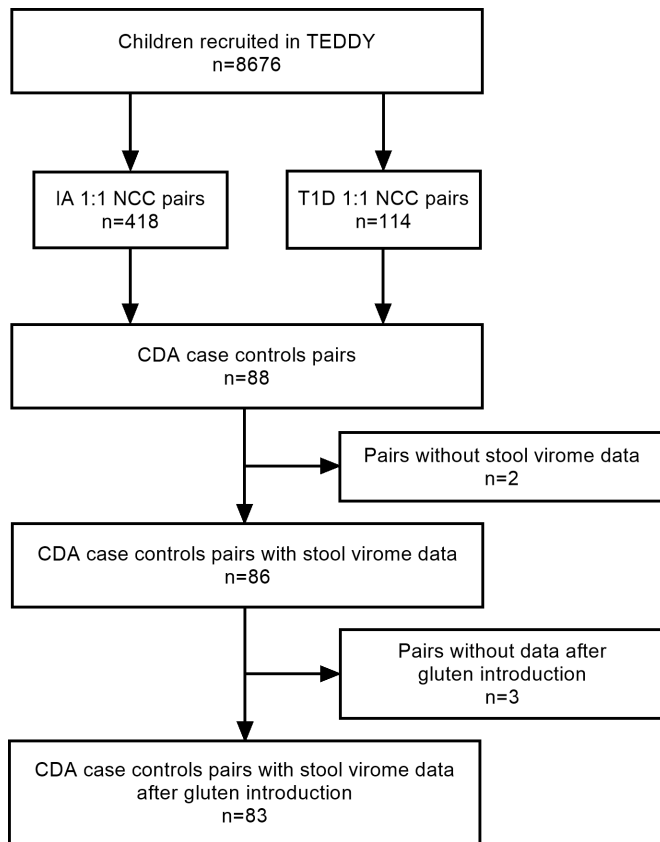


Figure 1 Flow chart describing the selection of the CDA case–control pairs for the study. CDA, coeliac disease autoimmunity; IA, islet autoantibody; NCC, nested case–control; T1D, type 1 diabetes.

serum tTG autoantibodies, is indicative of ongoing gluten-induced inflammatory response and may precede small bowel mucosal damage.^{3,4}

Higher gluten intake increases the risk of coeliac disease.^{5,6} However, it is not clear why not all genetically predisposed HLA-DQ2 and/or DQ8 positive individuals who eat gluten develop the disease. As a result, reasons leading to breakage of oral tolerance for gluten in coeliac disease remain unresolved. The possible role of infections in the development of coeliac disease has been supported in previous observational studies.^{7–10} These investigations have been extended by other studies, pointing to a potential role of viral infections, particularly by adenovirus, enterovirus, rotavirus and reovirus, in the disease pathogenesis.^{10–15} However, these findings are mainly based on cross-sectional or experimental studies. Thus, prospective studies are warranted to clarify if a virus or a group of viruses are involved in the aetiology of coeliac disease. More importantly, studies are limited in investigating a possible interaction between viral infections and gluten intake.

This study investigated the viral exposures prior to development of CDA from serial stool samples collected from children followed in a prospective birth cohort at genetic risk for both type 1 diabetes (T1D) and coeliac disease. By examining next-generation sequencing (NGS) metagenomics data, we assessed if viruses detected in the stool are associated with CDA and whether this association involved a possible interaction with gluten intake.

MATERIALS AND METHODS

The nested case–control (NCC) study design

Following new-born screening for high-risk HLA-DR-DQ genotypes, The Environmental Determinants of Diabetes in the Young

Table 1 Demographic data of the 83 nested case and control pairs

	Cases	Controls
HLA genotype, n (%)		
DQ2/DQ2	29 (35)	11 (13)
DQ2/DQ8	39 (47)	36 (43)
DQ8/DQ8	11 (13)	20 (24)
DQ8/X	4 (5)	15 (19)
Other (ineligible)		1 (1)
Age at CDA, months	31 (23, 46)	NA
Developed CD during follow-up, n (%)	28 (34)	NA
With CD-FDR, n (%)	6 (7)	3 (4)
Breastfeeding		
Ever breastfed, n (%)	83 (100)	83 (100)
Breastfeeding stopped, months	8 (5, 11)	8 (4, 12)
Gluten		
Age at introduction, months	6 (5, 7)	6 (5, 7)
Total intake by 2 years of age (g)	8.0 (5.4, 11.0)	7.6 (5.0, 11.8)
IA positivity, n (%)	41 (49)	15 (18)
Prior to CDA, n (%)	31 (37)	NA

For continuous variables, the median (25th percentile, 75th percentile) is reported. CD, coeliac disease; CDA, coeliac disease autoimmunity; CD-FDR, subject having a first degree relative with coeliac disease; IA, islet autoantibody; NA, not applicable.

(TEDDY) enrolled 8676 children before 4.5 months of age for a 15-year follow-up study with the main aim of identifying genetic and environmental triggers associated with T1D and coeliac disease.¹⁶ Annual screening for CDA started at the age of 2 years by detection of tTG autoantibodies using radiobinding assays, as previously described.¹⁷ If a sample was tTG-autoantibody positive, all of the child's earlier available samples were tested to determine the age of seroconversion. CDA was defined as being positive for tTG autoantibodies in two consecutive samples at least 3 months apart.¹⁸

From this cohort, two nested case–control (NCC) studies were constructed to improve the efficiency of multiple biomarker studies, with one focused on islet autoimmunity (IA) and the other on T1D. Cases and controls were identified as of 31 May 2012, then all available samples meeting the design criteria by that time were processed in the laboratories chosen for each biomarker analysis. Case–control pairs were matched for T1D family history (defined as having a first-degree relative with T1D), gender and clinical site location in the region where the participant was enrolled.¹⁹ All children in the 1:1 NCC studies for gut virome analysis that had been screened for CDA were considered for the present study. Each case–control pair included a CDA positive child ('case') matched with a child ('control') who was CDA-free for at least 6 months from the CDA case's age of seroconversion.

Identified were 88 CDA case–control pairs. Among those, only the 83 pairs (44 were females and 39 were males) whose stool virome data were available after introducing gluten in their diet were included in the final analysis (figure 1, table 1). The distribution by country was as follows: USA (n=25), Finland (n=13), Germany (n=5) and Sweden (n=40). There were 16 pairs with family history of T1D. Of the CDA cases, 41 were confirmed for IA positivity and 31 of these developed IA prior to CDA. Of the controls, 11 were IA positive. During the follow-up, 28 of the CDA cases developed coeliac disease. Six CDA cases and three controls had a first-degree relative with coeliac disease. As for the rotavirus vaccination status, 15 CDA cases and 20 controls

were vaccinated. Other characteristics of the cases and controls are shown in table 1.

Detection of viral stool sequences

Stool samples were collected monthly from 3 months until 2 years of age. NGS viral sequences were assayed in serial stool samples using a custom offline version of Vipe.²⁰ Vipe virus population profiling pipeline components include standard scripts for base quality, trimming, chimera detection while de novo assembled contigs were generated via integration of local assembly methods SPAdes and Velvet.^{21 22} These contigs were mapped to NCBI virus database using BLAST²³ resulting in a sample-based general virus population profile. For the present study, we focused on enterovirus, adenovirus, astrovirus, norovirus, reovirus and rotavirus. To approximate serotype and increase specificity, viral structural capsid-specific remapping on positive NGS samples was performed for enteroviruses and adenoviruses. For this analysis, MAPQ alignment cut-off of 20, representing greater than 0.99 probability was applied. The capsid resource contains Genbank and Tampere Virology Group-selected strains of adenovirus hexon, fibre and penton regions as well as enterovirus strains *Coxsackievirus A*, *Coxsackievirus B* and selected *Echovirus P1* protein (VP1-4) regions. This study included 1507 samples processed, after introducing gluten in diet.

Dietary data from food records

By 2 years of age, information on breastfeeding and the timing of introduction to gluten-containing cereals were collected from validated questionnaires at each clinic visit occurring every 3 months.⁵ Information on gluten consumption was collected at each clinic visit every 3 months up to 1 year of age and biannually thereafter (24 hours recall at 3-month visit and subsequently 3-day food records). Amount of gluten intake was calculated by multiplying the amount of vegetable protein in gluten-containing flours by a factor of 0.8.²⁴ From the 3-day food records, daily consumption (g/day) was obtained as the mean of 3 days of consumption.

Statistical analysis

The case-control pairs identified from the TEDDY NCC design were used to examine whether viral profiles differed by the CDA status. Conditional logistic regression was used to compare the cumulative appearance of viral exposures, after adjusting for HLA. Viral exposures were categorised by age: <1 year of life and from 1 to <2 years of life. From pairs where the case seroconverted prior to 2 years of age, only the samples available prior to age of seroconversion were included in the analysis. Interaction with cumulative gluten intake on the risk of CDA was examined. The cumulative gluten intake was obtained from the sum of daily consumption (g/day) from all clinical visits by 2 years of age. None reported any gluten consumption at 3-month visit. Additionally, the effects of the enterovirus sequence reads from 1 to <2 years of life were assessed in three groups by the total gluten intake: low (<33rd percentile), middle (33–66rd percentile) and high (>66rd percentile) based on unique individuals included in the analysis. Two-sided p-values are reported. Statistical significance was determined when the p-value was <0.05. All statistical analyses were performed using SAS V.9.4.

RESULTS

Association of viral exposures with CDA

The highest coverage of stool samples (available in 72.9% of the case-control pairs) was at 9 months where after the number gradually declined and at 24 months samples were available in 21.7% of the pairs (figure 2). Among the available stool samples at each collection age, the percentage of samples positive for any virus fluctuated between 22% and 50% without no obvious peaks at any collection age (figure 2A). The frequency of enterovirus positive samples ranged from 0% to 21% from the age 6 months onwards (figure 2B).

Between the time of first introduction of gluten at median 6 months of age and 1 year of age, 63 cases compared with 72 controls had at least one viral exposure (table 2). Enterovirus and adenoviruses were detected in 17 and 56 cases compared with 19 and 65 controls, respectively. Reoviral exposures were detected in only one case and one control, and rotaviral exposures were detected only in one of the controls. The cumulative

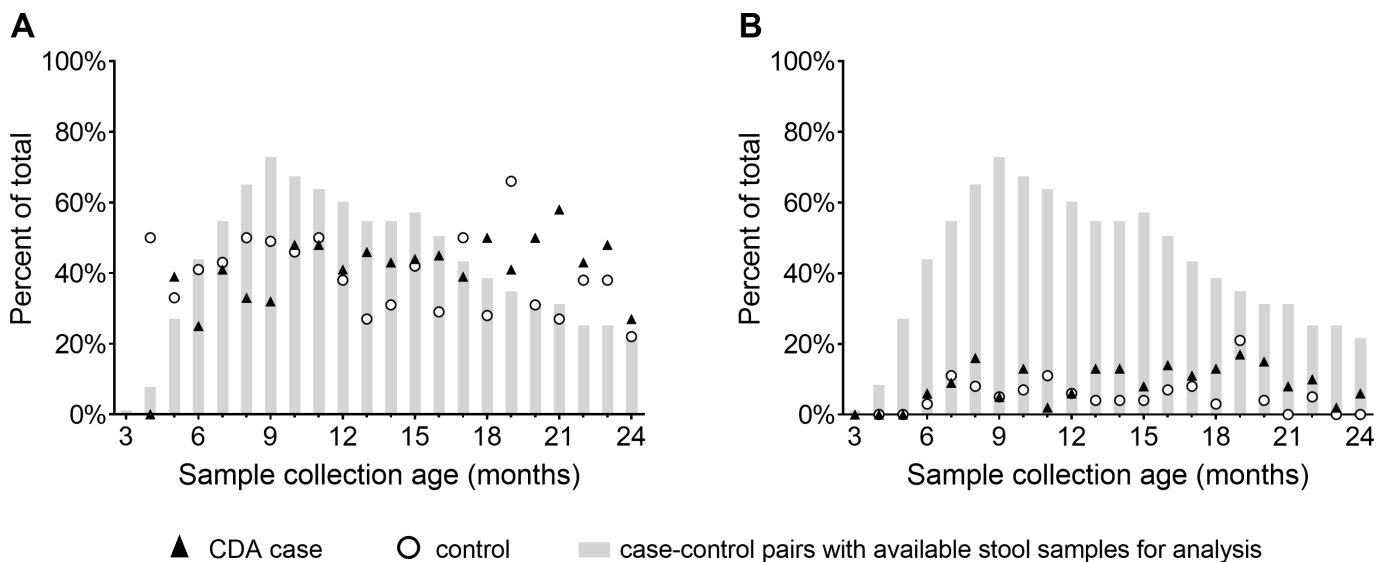


Figure 2 Stool samples positive for (A) any of the investigated viruses and (B) enteroviruses by 2 years of age as a percentage of samples available at each collection age. Filled triangles denote cases with CDA and unfilled circles controls. Bars represent the percentage of case-control pairs from whom stool samples were available for analysis at each collection age. CDA, coeliac disease autoimmunity.

Table 2 HLA-adjusted OR of cumulative virus detections in stool up to 1 year of age from a matched CDA case and control study

Virus/serotype	OR (95% CI)	P value	Cases positive (n)†	Controls positive (n)†
Any virus	0.68 (0.49 to 0.94)	0.02	63	72
HEV*	0.97 (0.56 to 1.70)	0.92	17	19
HEV A	1.65 (0.56 to 4.84)	0.36	8	6
CVA	6.69 (0.70 to 63.79)	0.1	8	2
HEV B	1.64 (0.77 to 3.53)	0.2	12	8
CVB	1.85 (0.81 to 4.24)	0.15	12	7
Echovirus	1.92 (0.68 to 5.45)	0.22	9	6
HAdV*	0.69 (0.48 to 0.99)	0.04	56	65
HAdV A	0.56 (0.13 to 2.36)	0.43	3	3
HAdV B	NA	NA	1	3
HAdV C	0.72 (0.37 to 1.37)	0.31	19	23
HAdV F	3.62 (0.70 to 18.88)	0.13	10	5
Astrovirus	0.94 (0.42 to 2.11)	0.89	15	18
Norovirus	1.10 (0.48 to 2.52)	0.82	12	11
Reovirus	0.71 (0.03 to 18.69)	0.84	1	1
Rotavirus	NA	NA	0	1

Total number of pairs with available stool samples is 79.

*Species and serotypes of HEV and HAdV are based on virome capsid mapping after Vipe genus taxonomy identification as follows: enterovirus capsid (VP1-4) repository includes CVA9, CVB1-6, Echovirus 6, 11, 18, 25 and 30. Adenovirus resource includes penton, hexon and fibre regions from prototypes AC_000007, X73487, JX423382, NC_010956, NC_001454 and KF303071.

†Number refers to the number of children having at least one viral sequence read in the stool.

CDA, coeliac disease autoantibody positivity; CVA, coxsackievirus A; CVB, coxsackievirus B; HAdV, human adenovirus; HEV, human enterovirus; NA, not applicable.

amount of any viral exposures (OR 0.68, 95% CI 0.49 to 0.94, $p=0.02$), and specifically those by adenoviruses (OR 0.69, 95% CI 0.48 to 0.99, $p=0.04$), were inversely associated with CDA when adjusting for HLA (table 2). None of the other individual viruses were associated with CDA.

Between 1 and 2 years of age, 59 cases compared with 58 controls had at least one positive stool sample with any of the selected viruses (table 3). Adenoviruses were detected in 52 cases compared with 55 controls, whereas enteroviruses were detected in 31 cases compared with 16 controls. Reovirus sequences were detected in one case, but none of the controls, while rotavirus sequences were not detected in any subjects. Cumulative number of positive stool samples for any virus was associated with an increased risk for CDA (OR 1.60, 95% CI 1.12 to 2.29, $p=0.01$). Of the different viruses, enteroviruses conferred the strongest positive association with CDA (OR 2.56, 95% CI 1.19 to 5.51, $p=0.02$) (table 3). When excluding the enteroviruses, the association of any viruses with the development of CDA was lost (OR 1.35, 95% CI 0.92 to 1.98, $p=0.13$). The ORs for the enterovirus B group were in the same direction when excluding the 42 pairs where either the case or control developed IA or T1D prior to seroconversion of CDA, although the reduced sample size did not have enough power to show differences between the groups (table 3).

Effects of feeding habits and viral exposures on the risk of CDA

The cumulative amount of any viral infections or enteroviruses during the period after gluten introduction but while breastfeeding was still ongoing was not associated with CDA (OR 1.15, 95% CI 0.72 to 1.82, $p=0.56$ and OR 0.98, 95% CI 0.43 to 2.21, $p=0.96$, respectively). When restricting the analysis to the time period after the end of any breastfeeding and to stool

Table 3 HLA-adjusted OR of cumulative viral detections in stool between 1 and 2 years of age: a matched CDA case and controls study

Virus/serotype	OR (95% CI)	P value	Cases positive (n)†	Controls positive (n)†	OR (95% CI) Excluding pairs with IA first‡
Any virus	1.60 (1.12 to 2.29)	0.01	59	58	1.02 (0.56 to 1.83)
HEV*	2.56 (1.19 to 5.51)	0.02	31	16	1.04 (0.42 to 2.56)
HEV A	2.10 (0.67 to 6.60)	0.2	13	8	0.65 (0.13 to 3.15)
CVA	2.53 (0.73 to 8.78)	0.15	7	5	0.64 (0.06 to 6.96)
HEV B	2.64 (0.84 to 8.36)	0.1	15	7	1.32 (0.32 to 5.41)
CVB	6.00 (1.27 to 28.46)	0.02	16	6	2.45 (0.45 to 13.3)
Echovirus	2.27 (0.68 to 7.61)	0.18	11	6	1.86 (0.42 to 8.21)
HAdV*	1.41 (0.99 to 2.02)	0.05	52	55	1.03 (0.59 to 1.77)
HAdV A	5.11 (0.71 to 36.63)	0.1	9	1	NA
HAdV B	6.12 (0.55 to 67.70)	0.14	4	1	2.46 (0.21 to 28.7)
HAdV C	1.74 (0.82 to 3.70)	0.15	19	17	1.31 (0.45 to 3.88)
HAdV F	1.09 (0.23 to 5.11)	0.91	7	7	2.87 (0.21 to 39.1)
Astrovirus	0.70 (0.29 to 1.68)	0.42	13	18	0.45 (0.10 to 1.99)
Norovirus	1.04 (0.42 to 2.62)	0.93	10	11	0.33 (0.05 to 3.00)
Reovirus	NA	NA	1	0	NA
Rotavirus	NA	NA	0	0	NA

Total number of pairs with available stool samples is 72.

*Species and serotypes of HEV and HAdV are based on virome capsid mapping after Vipe genus taxonomy identification as follows: enterovirus capsid (VP1-4) repository includes CVA9, CVB1-6, Echovirus 6, 11, 18, 25 and 30. Adenovirus resource includes penton, hexon and fibre regions from prototypes AC_000007, X73487, JX423382, NC_010956, NC_001454 and KF303071.

†Number refers to the number of children having at least one viral sequence read in the stool.

‡Excluded pairs were those with CDA cases being positive for IA or T1D prior to CDA ($n=31$) or controls being IA positive by the CDA cases' seroconversion ($n=11$).

CDA, coeliac disease autoantibody positivity; CVA, coxsackievirus A; CVB, coxsackievirus B; HAdV, human adenovirus; HEV, human enterovirus; IA, islet autoimmunity; NA, not applicable; T1D, type 1 diabetes.

Table 4 HLA-adjusted interactions of specific viruses or virus serotypes between 1 and 2 years of age with the cumulative amount of ingested gluten up to the age of 2 years

Virus/serotype	P value
Any virus	0.41
HEV	0.03
HEV A	0.05
CVA	0.83
HEV B	0.04
CVB	0.1
Echovirus	0.1
HAdV	0.5
HAdV A	0.81
HAdV B	0.44
HAdV C	0.76
HAdV F	0.15
Astrovirus	0.37
Norovirus	0.08
Reovirus	NA
Rotavirus	NA

CVA, coxsackievirus A; CVB, coxsackievirus B; HAdV, human adenovirus; HEV, human enterovirus; NA, not applicable.

samples collected between 1 and 2 years of age when the children were exposed to gluten, both the cumulative amount of any virus sequence reads (OR 1.41, 95% CI 1.00 to 2.00, $p=0.05$) and enterovirus sequence reads (OR 2.47, 95% CI, 1.12 to 5.48, $p=0.03$) were associated with CDA.

There was a significant interaction between enteroviruses between 1 and 2 years of age and the cumulative gluten intake by 2 years of age in the risk of CDA ($p=0.03$) (table 4). The risk of CDA was the highest among enterovirus positive children who had the highest cumulative gluten intake by 2 years of age (OR 8.3 (95% CI 1.8 to 37.1), as compared with those consuming middle or low amounts (OR 2.9, 95% CI 1.2 to 7.1 and OR 1.0, 95% CI 0.4 to 2.8, respectively) figure 3.

DISCUSSION

This study showed that the cumulative number of stool enteroviral exposures between 1 and 2 years of age was associated with CDA. In addition, an interaction between enteroviral exposures and gluten intake was observed. More importantly, the risk of CDA was increased in cases reporting higher intake of gluten. These results indicate that enteroviral exposure augmented by

higher gluten intake could act as triggers of coeliac disease in genetically at-risk children.

Only few previous studies have reported an association between enteroviruses and coeliac disease of which one study found an increased number of tTG autoantibody positive subjects among individuals with a proven enteroviral infection.²⁵ In addition, enteroviruses have previously been detected in the small bowel mucosa of coeliac disease patients.²⁶ Moreover, enteroviral infections prior to 1 year of age were not associated with increased risk of coeliac disease, while those occurring after the age of 1 year increased the risk.^{12 27} Our findings are consistent with both of these previous studies. Additionally, in line with the Norwegian study,¹² our results also showed that enteroviruses increased the risk of CDA only when restricting the analysis to the period when breastfeeding had ceased, but not while breastfeeding was still ongoing. As breastfeeding does not seem to be associated with the development of coeliac disease,^{28 29} this finding might indicate that enteroviral infections modulate CDA risk at an age when the child has ceased being breastfed.

This study extends previous investigations by showing an interaction between cumulative enteroviral exposures between age 1 and 2 with cumulative amount of gluten intake by 2 years of age. In addition, enteroviruses were associated with high risk, particularly among children reporting higher gluten intake, indicating that the gluten amount could amplify the effect of enteroviruses in the development of CDA. As comprehensive BLAST searches of the enterovirus contigs compiled from the NGS-derived enterovirus sequence reads did not reveal hits to gluten (data not shown), molecular mimicry likely does not account for the additive risk effect. Distinct isolate of reovirus, T1L has recently been shown to abrogate oral tolerance to gluten in a mechanism involving a type 1 interferon-induced activation of gluten-specific inflammatory T cells and inhibition of regulatory T cells.^{15 30} Along with these, the T1L reovirus infection was also reported to activate tTG.¹⁵ As enteroviruses have been shown to induce type 1 interferons,³¹ it could be speculated that enteroviruses may promote the development of coeliac disease by a mechanism involving an enterovirus-induced type 1 interferon-mediated breakage of oral tolerance to gluten coupled with activation of tTG similar to reovirus. Higher gluten intake will result in more gliadin peptides available for tTG-mediated deamidation resulting in augmented activation of the immune system ultimately leading chronic gut inflammation.

According to our results, cumulative adenoviral infections were associated with reduced risk for CDA after gluten introduction up to 1 year of age suggesting a putatively protective role.

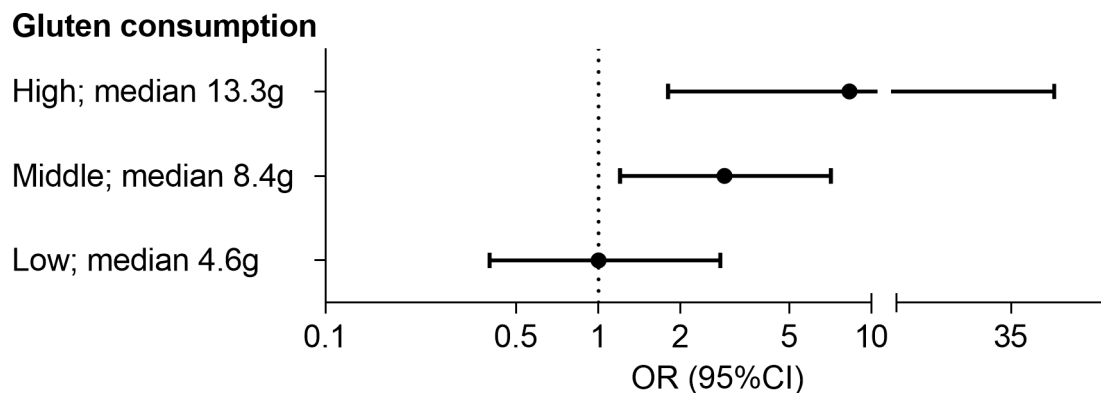


Figure 3 Effect of the enteroviral exposures between 1 and 2 years of age and risk of coeliac disease autoimmunity stratified by cumulative gluten consumption up to 2 years of age.

Our finding thus contradicts previous results that have pointed to either an inductive role^{11 32 33} or no role at all.^{12 34 35} Since none of the previous studies focused on the age window before 1 year of age, our finding of a possible protective effect of adenoviral infections warrants further research. However, if additional studies confirm the protective effect of adenovirus infections prior to 1 year of age, the diverging findings of negative and positive associations before and after 1 year of age raises the idea of a more generalised effective time-window effect. Contributing factors might involve the loss of immune protection from maternal antibodies as well as changes, also other than breast-feeding in the feeding habit that occur around 1 year of age.

Previous studies based on vaccination data or determination of rotavirus targeted antibodies have suggested that rotaviruses could trigger coeliac disease.^{10 13 14 36} Other studies applying animal models or circulating reovirus antibodies indicate a role of reoviruses in the disease pathogenesis.^{15 30} In the present study, rotavirus and reovirus sequence reads were detected in a minority of the stool samples. In the case of rotavirus, this is consistent with previous reports using PCR for virus detection in the stool.³⁷ Both rotavirus and reovirus are rapidly cleared from the stools after an infection.³⁸ Thus, more frequent stool sampling and other methods would therefore be required to detect short-lived or transiently occurring stool viruses in this study.

A major strength of this study is the design of analysing viral exposures from prospectively collected stool samples from birth until 2 years of age in cases compared with matched controls. Another strength is that our analysis focuses on viral exposures detected using capsid protein mappings prior to development of CDA, which can delineate disease progression and timing on the study outcome. However, the CDA case-control pairs were identified among either IA or T1D NCC pairs which resulted in high percentage of IA positive subjects in our cohort. As enterovirus infections are associated with the induction of beta-cell autoimmunity,³⁹ it is possible that the higher frequency of IA positive subjects among the CDA cases may have confounded the results. In addition, the variation of sequence coverage and number of subject pairs restricted the identification of individual enterovirus types, which limits the analysis to species level. Another potential limitation was not adjusting for having a first-degree family member with coeliac disease. This was not carried out in order to avoid a selection bias due to more likely antibody screening of family members among the CDA cases.

In conclusion, the present study found that a cumulative number of stool enteroviral exposures is associated with CDA in children at genetic risk for coeliac disease. In addition, the interaction of enteroviral exposures and higher gluten intake indicate a cumulative effect of these factors in the development of CDA.

Author affiliations

¹Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland

²Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland

³Morsani College of Medicine, University of South Florida, Tampa, Florida, USA

⁴Center for Child Health Research, Tampere University and Tampere University Hospital, Tampere, Finland

⁵The University Consortium of Seinäjoki, Seinäjoki, Finland

⁶University of Colorado Denver, Anschutz Medical Campus, Aurora, Colorado, USA

⁷Digestive Health Institute, Children's Hospital Colorado, Aurora, United States

⁸Ludwig-Maximilians-Universität München, München, Bayern, Germany

⁹Division of Paediatric Gastroenterology and Hepatology, Dr von Hauner Children's Hospital, München, Germany

¹⁰Barbara Davis Center for Childhood Diabetes, University of Colorado Denver, Denver, Colorado, USA

¹¹Pacific Northwest Research Institute, Seattle, Washington, USA

¹²Research Centre for Integrative Physiology and Pharmacology, Institute of Biomedicine, University of Turku, Turku, Finland

¹³Department of Paediatrics, Turku University Hospital, Turku, Finland

¹⁴Klinikum Rechts der Isar, Technische Universität München, München, Bayern, Germany

¹⁵Institute of Diabetes Research, Helmholtz Zentrum München, Germany

¹⁶Forschergruppe Diabetes e.V., Neuherberg, Germany

¹⁷National Institute of Diabetes and Digestive and Kidney Disease, National Institutes of Health, Bethesda, Maryland, USA

¹⁸Alkek Center for Metagenomics and Microbiome Research, Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, Texas, USA

¹⁹The Diabetes and Celiac Disease Unit, Department of Clinical Sciences, Lund University, Lund, Sweden

Collaborators Aaron Barbour; Kimberly Bautista; Judith Baxter; Daniel Felipe-Morales; Kimberly Driscoll; Brigitte I Frohnert; Marisa Stahl; Patricia Gesualdo; Michelle Hoffman; Rachel Karban; Jill Norris; Stesha Peacock; Hanan Shorosh; Andrea Steck; Megan Stern; Erica Villegas; Kathleen Waugh; Olli G Simell; Annika Adamsson; Suvi Ahonen; Mari Åkerlund; Leena Hakola; Anne Hekkala; Henna Holappa; Anni Ikonen; Jorma Ilonen; Sinikka Jäminki; Sanna Jokipuu; Leena Karlsson; Jukka Kero; Miia Kähönen; Mikael Knip; Minna-Liisa Koivikko; Merja Koskinen; Mirva Koreasalo; Jarita Kytölä; Tiina Latva-aho; Maria Lönnrot; Elina Mäntymäki; Markus Mattila; Maija Miettinen; Katja Multasuo; Teija Mykkänen; Tiina Niininen; Sari Niinistö; Mia Nyblom; Sami Oikarinen; Paula Ollikainen; Zhian Othmani; Sirpa Pohjola; Petra Rajala; Jenna Rautanen; Eino Riikonen; Eija Riski; Miia Pekkola; Minna Romo; Satu Ruohonen; Satu Simell; Maija Sjöberg; Aino Stenius; Päivi Tossavainen; Mari Vähä-Mäkilä; Sini Vainionpää; Eeva Varjonen; Riitta Veijola; Irene Viinikangas; Suvi M Virtanen; Jin-Xiong She; Desmond Schatz; Diane Hopkins; Leigh Steed; Jennifer Bryant; Katherine Silvis; Michael Haller; Melissa Gardiner; Richard McIndoe; Ashok Sharma; Stephen W Anderson; Laura Jacobsen; John Marks; Ezio Bonifacio; Anita Gavrisan; Cigdem Gezginci; Anja Heublein; Verena Hoffmann; Sandra Hummel; Andrea Keimer; Annette Knopff; Charlotte Koch; Claudia Ramminger; Roswith Roth; Marlon Scholz; Joanna Stock; Katharina Warncke; Lorena Wendel; Christiane Winkler; Åke Lernmark; Carin Andrén Aronsson; Maria Ask; Rasmus Bennet; Corrado Cilio; Helene Engqvist; Emelie Ericson-Hallström; Annika Fors; Lina Fransson; Thomas Gard; Monika Hansen; Hanna Jisser; Fredrik Johansen; Berglind Jonsdottir; Silvija Jovic; Helena Elding Larsson; Marielle Lindström; Markus Lundgren; Marlena Maziarz; Maria Månsson-Martinez; Maria Markan; Jessica Melin; Zeliha Mestan; Caroline Nilsson; Karin Ottosson; Kobra Rahmati; Anita Ramelius; Falastin Salami; Anette Sjöberg; Birgitta Sjöberg; Malin Svensson; Carina Törn; Anne Wallin; Åsa Wimar; Sofie Åberg; Michael Killian; Claire Cowen Crouch; Jennifer Skidmore; Masumeh Chavoshi; Rachel Hervey; Rachel Lyons; Arlene Meyer; Denise Mulenga; Jared Radtke; Matei Romancik; Davey Schmitt; Sarah Zink; Dorothy Becker; Margaret Francisus; MaryEllen Dalmagro-Elias Smith; Ashi Daftary; Mary Beth Klein; Chrystal Yates; Sarah Austin-Gonzalez; Maryouri Avendano; Sandra Baethke; Rasheedah Brown; Brant Burkhardt; Martha Butterworth; Joanna Clasen; David Cuthbertson; Stephen Danky; Christopher Eberhard; Steven Fiske; Jennifer Garmeson; Veena Gowda; Kathleen Heyman; Belinda Hsiao; Christina Karges; Francisco Perez Laras; Qian Li; Shu Liu; Xiang Liu; Kristian Lynch; Colleen Maguire; Jamie Malloy; Cristina McCarthy; Aubrie Merrell; Hemang Parikh; Ryan Quigley; Cassandra Remedios; Chris Shaffer; Laura Smith; Susan Smith; Noah Sulman; Roy Tamura; Dena Tewey; Michael Toth; Ulla Uusitalo; Kendra Vehik; Ponnai Vijayakandipan; Keith Wood; Jimin Yang; Michael Abbondandolo; Lori Ballard; David Hadley; Wendy McLeod; Steven Meulemans; Liping Yu; Dongmei Miao; Polly Bingley; Alistair Williams; Kyla Chandler; Olivia Ball; Ilana Kelland; Sian Grace; Masumeh Chavoshi; Jared Radtke; Sarah Zink; Nadim J Ajami; Matthew C Ross; Jacqueline L O'Brien; Diane S Hutchinson; Daniel P Smith; Matthew C Wong; Xianjun Tian; Tulin Ayvaz; Auriolo Tamegno; Nguyen Truong; Hannah Moreno; Lauren Riley; Eduardo Moreno; Tonya Bauch; Lenka Kusic; Ginger Metcalf; Donna Muzny; HarshaVardhan Doddapaneni; Richard Gibbs; Sandra Ke; Niveen Mulholland; Kasia Bourcier; Thomas Briesse; Suzanne Bennett Johnson; Eric Triplett.

Contributors Study concept and design: JL, HH, REL, MN, KK, EL, SK, MR, WH, JT, AGZ, BA, JPK, JFP, KL, DA; administrative, technical and material support: WH, JT, BA, JPK; acquisition, analysis and interpretation of the data: JL, HSL, HH, REL, MN, KK, EL, SK, WH, JT, AGZ, BA, JPK, JFP, KL, DA; statistical analysis: JL, HSL, MN, JPK, JFP; drafting of the manuscript: JL, HH, KL, DA; obtaining funding: MR, WH, JT, AGZ, BA, JPK; study supervision: HH, MR, WH, JT, AGZ, BA, JPK, DA. All authors have participated in the critical revision of the manuscript for important intellectual content.

Funding The TEDDY Study was funded by grants U01 DK63829, U01 6 DK63861, U01 DK63821, U01 42 DK63865, U01 DK63863, U01 DK63836, 7 U01 DK63790, UC4 DK63829, UC4 DK63861, UC4 43 DK63821, UC4 8 DK63865, UC4 DK63863, UC4 DK63836, UC4 DK95300, UC4 DK100238, and 44 UC4 DK106955, UC4 DK112243, UC4 DK117483, and Contract No. HHSN267200700014C from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of 2 Allergy and Infectious Diseases (NIAID), Eunice Kennedy Shriver National Institute of Child Health 3 and Human Development (NICHD), National Institute of Environmental Health Sciences (NIEHS), 4 Centers for Disease

Control and Prevention (CDC) and JDRF. This work supported in part by the 5 NIH/NCATS Clinical and Translational Science Awards to the University of Florida (UL1 TR000064) and the University of Colorado (UL1 TR001082).

Competing interests HH is a shareholder and chairman of the board of Vactech Ltd, which develops vaccines against picornaviruses.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement De-identified datasets generated and analysed during the current study will be made available by request from the NIDDK Central Repository at <https://www.niddkrepository.org/studies/teddy>. The TEDDY metagenomics next-generation sequencing (NGS) data that support the findings of this study will be available by request from NCBI's database of Genotypes and Phenotypes (dbGaP) with the primary accession code phs001442.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Katri Lindfors <http://orcid.org/0000-0001-7417-5151>

Kalle Kurppa <http://orcid.org/0000-0003-4757-2164>

Daniel Agardh <http://orcid.org/0000-0003-1435-1234>

REFERENCES

- Vojdani A. A potential link between environmental triggers and autoimmunity. *Autoimmune Dis* 2014;2014:1–18.
- Dieterich W, Ehnis T, Bauer M, et al. Identification of tissue transglutaminase as the autoantigen of celiac disease. *Nat Med* 1997;3:797–801.
- Kurppa K, Collin P, Viljamaa M, et al. Diagnosing mild enteropathy celiac disease: a randomized, controlled clinical study. *Gastroenterology* 2009;136:816–23.
- Kurppa K, Ashorn M, Iltanen S, et al. Celiac disease without villous atrophy in children: a prospective study. *J Pediatr* 2010;157:373–80.
- Andrén Aronsson C, Lee H-S, Koletzko S, et al. Effects of gluten intake on risk of celiac disease: a case-control study on a Swedish birth cohort. *Clin Gastroenterol Hepatol* 2016;14:403–9.
- Mårild K, Dong F, Lund-Blix NA, et al. Gluten intake and risk of celiac disease: long-term follow-up of an at-risk birth cohort. *Am J Gastroenterol* 2019.
- Myléus A, Hernel O, Gothefors L, et al. Early infections are associated with increased risk for celiac disease: an incident case-referent study. *BMC Pediatr* 2012;12:194.
- Canova C, Zabeo V, Pitter G, et al. Association of maternal education, early infections, and antibiotic use with celiac disease: a population-based birth cohort study in northeastern Italy. *Am J Epidemiol* 2014;180:76–85.
- Mårild K, Kahrs CR, Tapia G, et al. Infections and risk of celiac disease in childhood: a prospective nationwide cohort study. *Am J Gastroenterol* 2015;110:1475–84.
- Kemppainen KM, Lynch KF, Liu E, et al. Factors that increase risk of celiac disease autoimmunity after a gastrointestinal infection in early life. *Clin Gastroenterol Hepatol* 2017;15:694–702.
- Kagnoff MF, Paterson YJ, Kumar PJ, et al. Evidence for the role of a human intestinal adenovirus in the pathogenesis of coeliac disease. *Gut* 1987;28:995–1001.
- Kahrs CR, Chuda K, Tapia G, et al. Enterovirus as trigger of coeliac disease: nested case-control study within prospective birth cohort. *BMJ* 2019;34.
- Stene LC, Honeyman MC, Hoffenberg EJ, et al. Rotavirus infection frequency and risk of celiac disease autoimmunity in early childhood: a longitudinal study. *Am J Gastroenterol* 2006;101:2333–40.
- Dolcino M, Zanoni G, Bason C, et al. A subset of anti-rotavirus antibodies directed against the viral protein VP7 predicts the onset of celiac disease and induces typical features of the disease in the intestinal epithelial cell line T84. *Immunol Res* 2013;56:465–76.
- Bouziat R, Hinterleitner R, Brown JJ, et al. Reovirus infection triggers inflammatory responses to dietary antigens and development of celiac disease. *Science* 2017;356:44–50.
- Hagopian WA, Erlich H, Lernmark Åke, et al. The environmental determinants of diabetes in the young (TEDDY): genetic criteria and international diabetes risk screening of 421 000 infants. *Pediatr Diabetes* 2011;12:733–43.
- Vehik K, Fiske SW, Logan CD, et al. Methods, quality control and specimen management in an international multi-centre investigation of type 1 diabetes: TEDDY. *Diabetes Metab Res Rev* 2013;29:557–67.
- Liu E, Lee H-S, Aronsson CA, et al. Risk of pediatric celiac disease according to HLA haplotype and country. *N Engl J Med* 2014;371:42–9.
- Burkhardt BR, McLeod W, et al. Biomarker discovery study design for type 1 diabetes in the environmental determinants of diabetes in the young (TEDDY) study. *Diabetes Metab Res Rev* 2014;30:424–34.
- Lin J, Kramna L, Autio R, et al. Vipie: web pipeline for parallel characterization of viral populations from multiple NGS samples. *BMC Genomics* 2017;18:378.
- Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455–77.
- Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 2008;18:821–9.
- Altschul SF, Gish W, Miller W, et al. Basic local alignment search tool. *J Mol Biol* 1990;215:403–10.
- van Overbeek FM, Uil-Dieterman IGA, Mol IW, et al. The daily gluten intake in relatives of patients with coeliac disease compared with that of the general Dutch population. *Eur J Gastroenterol Hepatol* 1997;9:1097–9.
- Sarmiento L, Galvan JA, Cabrera-Rode E, et al. Type 1 diabetes associated and tissue transglutaminase autoantibodies in patients without type 1 diabetes and coeliac disease with confirmed viral infections. *J Med Virol* 2012;84:1049–53.
- Oikarinen M, Tauriainen S, Oikarinen S, et al. Type 1 diabetes is associated with enterovirus infection in gut mucosa. *Diabetes* 2012;61:687–91.
- Simre K, Uibo O, Peet A, et al. Early-life exposure to common virus infections did not differ between coeliac disease patients and controls. *Acta Paediatr* 2019;108:1709–16.
- Szajewska H, Shamir R, Chmielewska A, et al. Systematic review with meta-analysis: early infant feeding and coeliac disease - update 2015. *Aliment Pharmacol Ther* 2015;41:1038–54.
- Silano M, Agostoni C, Sanz Y, et al. Infant feeding and risk of developing celiac disease: a systematic review. *BMJ Open* 2016;6:e009163.
- Brown JJ, Jabri B, Dermody TS. A viral trigger for celiac disease. *PLoS Pathog* 2015;11:e1007181.
- Hämäläinen S, Nurminen N, Ahlfors H, et al. Coxsackievirus B1 reveals strain specific differences in plasmacytoid dendritic cell mediated immunogenicity. *J Med Virol* 2014;86:1412–20.
- Lähdeaho M-L, Lehtinen M, Rissa H-R, et al. Antipeptide antibodies to adenovirus E1B protein indicate enhanced risk of celiac disease and dermatitis herpetiformis. *Int Arch Allergy Immunol* 1993;101:272–6.
- Lähdeaho M-L, Parkkonen P, Reunala T, et al. Antibodies to E1b Protein-Derived Peptides of Enteric Adenovirus Type 40 Are Associated with Celiac Disease and Dermatitis Herpetiformis. *Clin Immunol Immunopathol* 1993;69:300–5.
- Mahon J, Blair GE, Wood GM, et al. Is persistent adenovirus 12 infection involved in coeliac disease? A search for viral DNA using the polymerase chain reaction. *Gut* 1991;32:1114–6.
- Lawler M, Humphries P, O'Farrelly C, et al. Adenovirus 12 E1A gene detection by polymerase chain reaction in both the normal and coeliac duodenum. *Gut* 1994;35:1226–32.
- Hemming-Harlow M, Lähdeaho M-L, Mäki M, et al. Rotavirus vaccination does not increase type 1 diabetes and may decrease celiac disease in children and adolescents. *Pediatr Infect Dis J* 2019;38:539–41.
- Fan Y-M, Oikarinen S, Lehto K-M, et al. High prevalence of selected viruses and parasites and their predictors in Malawian children. *Epidemiol Infect* 2019;147:e90.
- Dermody TS, Parker JSL, Sherry B. Orthoreovirus. In: Knipe DM, Howley PM, eds. *Fields virology*. 6th ed. Philadelphia, PA: Lippincott, Williams, & Wilkins, 2013: Volume 2. 1304–46.
- Salminen K, Sadeharju K, Lönnrot M, et al. Enterovirus infections are associated with the induction of β -cell autoimmunity in a prospective birth cohort study. *J Med Virol* 2003;69:91–8.