Science of the Total Environment 768 (2021) 144627

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

An empirical test of the biodiversity hypothesis: Exposure to plant diversity is associated with a reduced risk of childhood acute lymphoblastic leukemia

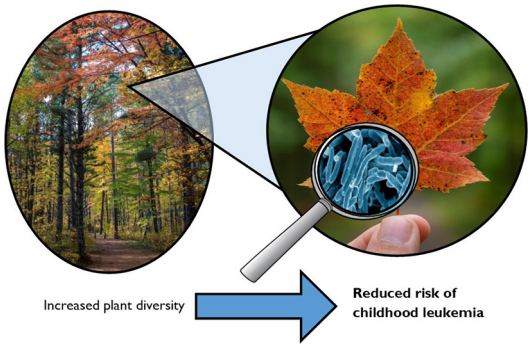
Geoffrey H. Donovan [a,b,*], Demetrios Gatziolis [b], Andrea 't Mannetje [a], Robert Weinkove [c,d,e], Caroline Fyfe [a], Jeroen Douwes [a]

[a] Center for Public Health Research, Massey University-Wellington Campus, PO Box 756, Wellington 6140, New Zealand
[b] USDA Forest Service, PNW Research Station, 620 SW Main, Suite 502, Portland, OR 97205, USA
[c] Cancer Immunotherapy Programme, Malaghan Institute of Medical Research, PO Box 7060, Wellington, New Zealand
[d] Wellington Blood & Cancer Centre, Capital & Coast District Health Board, Private Bag 7902, Wellington, New Zealand
[e] Department of Pathology & Molecular Medicine, University of Otago Wellington, Newtown, Wellington, New Zealand

HIGHLIGHTS

- Acute lymphoblastic leukemia (ALL) is the world's most common pediatric cancer.
- Exposure to plant diversity is associated with a reduced risk of ALL in a dose-dependent fashion.
- The protective effect of plant diversity may be mediated by changes in the microbiome.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 21 August 2020
Received in revised form 15 December 2020
Accepted 18 December 2020
Available online 10 January 2021

Editor: SCOTT SHERIDAN

Keywords:
Hygiene hypothesis
Cancer
Natural environment
Immune diseases
Greenness

ABSTRACT

The biodiversity hypothesis posits that declining biodiversity may be responsible, at least in part, for the global increase in immune diseases. However, few studies have been able to demonstrate a link between exposure to biodiversity and specific health outcomes. We test whether exposure to plant diversity protects against childhood acute lymphoblastic leukemia (ALL) by promoting immune maturation. Our sample consisted of all children born in New Zealand from 1998 to 2013 (n = 899,126; 264 ALL cases), which we followed from birth to age five. We calculated plant-diversity metrics using the Global Biodiversity Information Facility, which contains over two million geocoded plant records in New Zealand. Consistent with previous research, children who had always lived in an urban area, or who had an older mother, were at greater risk for ALL, whereas children with older siblings were at lower risk. In addition, we found that plant-diversity metrics based on the maximum number of plant genera a child was exposed to during the first two years of life were protective of ALL. Specifically, exposure to the highest tertile of plant diversity was associated with a reduction in ALL risk of 35% (95% CI: 11%–53%). Exposure to plant diversity, and associated microbial communities, may be a viable public-health intervention to reduce the risk of ALL and possibly other immune diseases.

Published by Elsevier B.V.

* Corresponding author at: USDA Forest Service, PNW Research Station, 620 SW Main, Suite 502, Portland, OR 97205, USA.
E-mail addresses: geoffrey.donovan@usda.gov (G.H. Donovan), demetrios.gatziolis@usda.gov (D. Gatziolis), A.Mannetje@massey.ac.nz (A. 't Mannetje), rweinkove@malaghan.org.nz (R. Weinkove), fyfca725@student.otago.ac.nz (C. Fyfe), J.Douwes@massey.ac.nz (J. Douwes).

https://doi.org/10.1016/j.scitotenv.2020.144627
0048-9697/Published by Elsevier B.V.

1. Introduction

The biodiversity hypothesis posits that reduced contact with the natural environment has adversely affected the human microbiota and its role in host immune regulation (Haahtela et al., 2013). It is based on a hypothesized link between two trends: the global loss of biodiversity and the concurrent increase in immune diseases especially in high-income countries (Isolauri 2001). Despite the biodiversity hypothesis receiving considerable attention in the scientific literature and the popular press, very few studies have demonstrated a direct link between exposure to biodiversity and specific health outcomes. We address this gap in the literature by testing the hypothesis that exposure to plant diversity protects against childhood acute lymphoblastic leukemia (ALL) in a large New Zealand birth cohort.

ALL is a malignancy of immune cell precursors, and the world's most common pediatric cancer, affecting 1 in 2000 children. Incidence is highest in high-income countries and is rising by approximately 1% per year (Linabery and Ross, 2008). Although contemporary chemotherapies cure most children with ALL, toxicities are substantial, and survivors can experience lifelong health problems (Inaba et al. 2013).

The most common form of childhood ALL, precursor B-cell ALL, is believed to develop in two discrete steps. First, an *in utero* genetic change generates a pre-leukemic clone. Second, after birth, a small proportion (<1%) of susceptible children develop a secondary genetic change following an infection, resulting in overt leukemia (Mori et al. 2002). Paradoxically, although infection is the proximate cause, microbial exposure in early life, a critical step in immune maturation, may reduce the risk of a dysregulated immune response to infection that triggers ALL development (Greaves 2018). Although direct evidence that microbial exposure protects against ALL is lacking, epidemiological studies consistently show that day care-attendance (Barrington-Trimis et al. 2017) and having older siblings (Ajrouche et al. 2015) (both associated with higher microbial exposure) are protective, whereas Caesarian delivery (lower microbial exposure) is a risk factor (Wang et al. 2017). Indeed, a recent review of the causal mechanisms of ALL (Greaves 2018) noted that through early-life microbial exposure "most cases of childhood ALL are potentially preventable".

1.1. Literature review

A necessary condition for diverse microbial exposure is a diverse microbial habitat, an important source of which is plants. There are 391,000 known vascular plant species with a total leaf area of 1,017,260,200 km^2 (twice the world's land area) supporting ~10^{26} bacterial cells (Vorholt 2012). Several studies have demonstrated that more diverse plant communities support greater microbial diversity. Notably, plants of the same species have similar microbial communities even when growing in different locations, whereas different plant species growing in the same location can support distinctly different microbes (Kowalchuk et al., 2002). Plant microbial communities are also an important source of airborne microbes. For example, a study comparing plant microbes to airborne microbes found that plant microbial communities are indistinguishable from airborne microbes 50 m downwind of a plant, but upwind microbes are unique (Lymperopoulou et al., 2016).

Diverse plant communities are also associated with greater diversity of the human microbiota. Specifically, a Scandinavian study (Hanski et al. 2012) found that people who live in areas with higher plant diversity have more diverse skin bacteria and a reduced risk of atopic sensitization. To our knowledge, no studies have shown a direct relationship between exposure to plant diversity and the diversity of the gut microbiota. However, several studies have found that animals who are exposed to more diverse plants have a more diverse gut microbiota. For example, Phillips et al. (2018) found significant differences in the diversity of the gut microbiota of urban and rural white-crowned sparrows, which the authors attribute to differences between urban and rural vegetation. More generally, a recent Chinese study demonstrated that geographical location is strongly associated with diversity of the gut microbiota (He et al. 2018).

Higher plant diversity is also associated increased diversity of the rhizosphere (Carney and Matson 2006), and emerging evidence suggests that the rhizosphere may influence the human microbiome. For example, Liddicoat et al. (2020) found that exposure to even trace amounts of dust from high-biodiversity soil influenced the composition of the gut microbiome of mice and reduced symptoms of anxiety. Therefore, the hypothesized protective effect of plant diversity on ALL may include both the atmosphere and the rhizosphere.

Before treatment, the gut microbiota of children with ALL is less diverse than that of healthy children, although it is unclear whether this is a cause or consequence of the disease (Rajagopala et al. 2016). In addition to ALL, dysbiosis of the gut microbiota is associated with a rapidly expanding range of adverse health outcomes (Cho and Blaser 2012).

Despite the fact that plants are a major source of microbial exposure, the association between exposure to plant diversity and ALL has not been studied, although a handful of papers have examined the relationship between exposure to plant diversity and other health outcomes. Most have focused on the relationship between self-reported wellbeing and plant-and-animal diversity, generally finding a positive association (Dallimer et al. 2012). Few studies have focused on specific health outcomes. One exception is a New Zealand study, which found that plant diversity is protective of childhood asthma (Donovan et al. 2018). Similarly, an Australian study (not focused exclusively on asthma) found that multiple biodiversity metrics were associated with lower rates of hospital admissions for respiratory disease (Liddicoat et al. 2018).

We test two hypotheses related to plant diversity and pediatric cancer. First, we hypothesize that children who are exposed to greater plant diversity are at a reduced risk of developing ALL. Second, we hypothesize that there is no association between exposure to plant diversity and the risk of developing other pediatric cancers.

2. Methods

2.1. Sample and study area

We followed all children born in New Zealand between January 1, 1998 and December 31, 2013 ($n = 946{,}098$) for five years. We were able to link 95.0% of these children to address and covariate data giving an analytical sample of 899,126. Data on outcomes and covariates were obtained using Statistics New Zealand's Integrated Data Infrastructure (IDI), which is a collection of linked individual-level databases that covers approximately 95% of New Zealand residents (Milne et al. 2019). We chose this sample frame, because residential history in the IDI is unreliable before 1998, and, because of a cancer reporting lag, 2013 was the latest year that allowed us to follow children for five years.

Our study was classified as out of scope by the New Zealand Health and Disabilities Ethics Committee, approved by Statistics New Zealand (approval #MAA2019-07), and conducted in accordance with national and international ethical guidelines for observational studies.

2.2. Outcomes

We identified cases of ALL (ICD 10 code: C91.00), and all other cancers, using the New Zealand cancer registry. Our sample contained 264 ALL cases by age 5 and 603 cases of other cancers.

2.3. Exposure metrics

The building blocks for our plant-diversity metrics were the number of plant species, genera, families, and orders a child was exposed to during each of the first five years of life (five annual values for each child). From these annual values, we created composite metrics. For example, for the first two years of life—the primary focus of this paper—we created three metrics from these annual values: mean exposure (mean

of the year one and year two values) as well as the maximum and minimum of these two annual values. We used mean, minimum, and maximum exposures in our analysis, because it is not clear whether immune maturation is a function of continuous microbial exposure, or if departures from mean lifetime exposure are more important. We focused on exposure during the first two years of a child's life, as multiple studies have shown that environmental exposures during this period play a critical role in immune maturation (Gollwitzer and Marsland 2015).

We used four data sources to create these annual exposure values (Fig. 1). First, we overlaid plant-occurrence data from the Global Biodiversity Information Facility (GBIF) with the 2008 New Zealand Land Cover Database (v4.1) (Land Information New Zealand 2012). GBIF contains >1 billion geocoded plant records of which 2,057,987 are from New Zealand. Most of these come from systematic surveys by government agencies, although some were collected by individuals participating in the iNaturalist Program ($n = 150{,}771$). The New Zealand Landcover Database classifies all land cover in mainland New Zealand into 1 of 33 classes (Schmitt et al., 2017).

We next calculated meshblock-level diversity by multiplying the number of species in each class by the proportion of a meshblock covered by that class and then summing across all land-cover classes (meshblocks are the smallest geographical unit at which Statistics New Zealand reports data). For example, if half a meshblock was covered by land-cover type 1 (1000 species nationally) and the other half by land-cover type 2 (500 species nationally), plant species diversity would be $((0{\cdot}5 * 1000) + (0{\cdot}5 * 500)) = 750$. We repeated this process for plant genus, family, and order. In New Zealand, the GBIF contains 9725 species, 2936 genera, 543 families, and 160 orders. Finally, we merged these data with residential history to calculate annual exposure values.

We used national totals of plant diversity in each land-cover class (rather than the number of each in a meshblock), because it is likely that sampling intensity is higher in areas where more people live or visit, and, in the presence of non-random sampling, this point-to-grid approach is most accurate, when points are aggregated to large units (such as national land-cover classes) (Schmitt et al., 2017). Therefore, our metrics should be viewed as the maximum plant diversity that a child was exposed in a given meshblock. Note that there were no plant records in 28,961 meshblocks, which account for 7.2% of the area of mainland New Zealand.

We calculated meshblock-level exposure metrics, because IDI provides residential history by meshblock. On average, 95 people live in a meshblock, and the mean area for urban and rural meshblocks is 20.4 and 2498 ha, respectively.

To measure intensity of plant exposure, we used the Normalized Difference Vegetation Index (NDVI), which is a greenness index (bounded by −1 and 1) derived from satellite imagery. Specifically, we used maximum annual NDVI derived from Landsat imagery archives (30 m resolution) (Chander et al. 2009) specific to each year of the analysis (annual values were calculated from multiple cloud-free scenes). As with plant diversity, we used these annual values to calculate exposure metrics for the first two year of a child's life. Fig. 2 shows plant genus diversity and NDVI for New Zealand and the capital, Wellington.

Maternal environmental exposure can influence a child's immune development (Jenmalm 2011). Therefore, we calculated prenatal exposure (conception to birth) following the same procedures we used for post-natal exposure (matching to maternal address). In addition, we classified each meshblock as urban or rural using definitions from the New Zealand Census (urban areas are defined as cities, towns and other conurbations of a thousand people or more) allowing us to calculate the proportion of a child's life spent living in urban and rural areas. Finally, we used Census data to calculate population density for each meshblock, as higher population density may be a risk factor for ALL (Hjalmars and Gustafsson 1999).

2.4. Covariates

We used the IDI to identify covariates that have previously been associated with ALL, including number of older siblings, mother's age, gender, ethnicity, and birth weight. To account for neighborhood-level social deprivation, we used NZDep, a well-validated index based on nine census

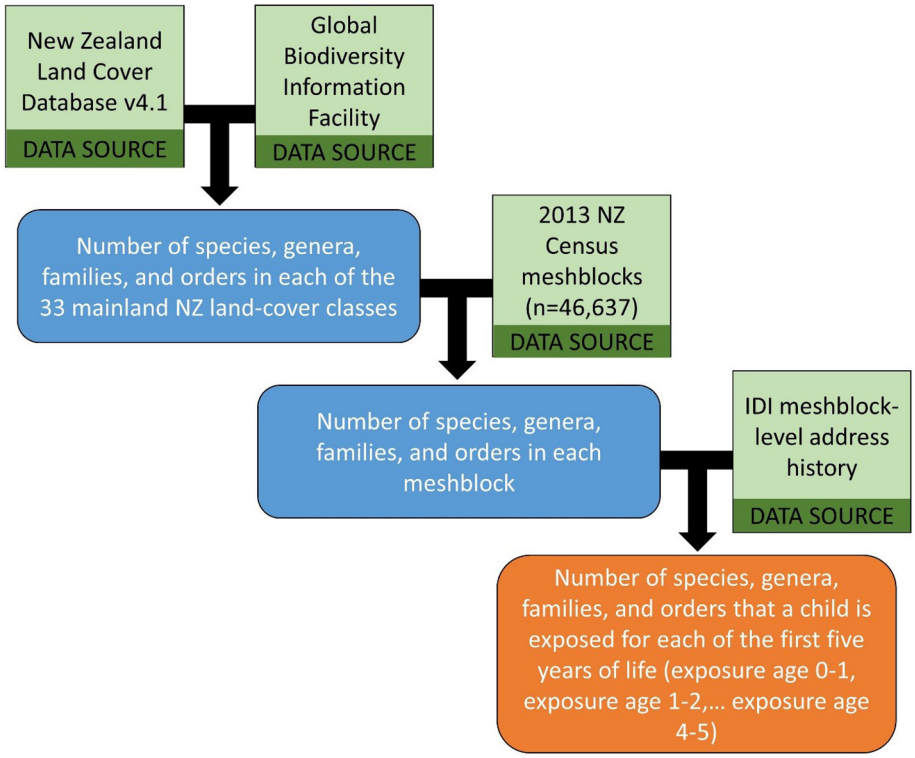


Fig. 1. Flowchart showing how annual plant-diversity exposures were calculated. Figure credit: Kendra Wendel.

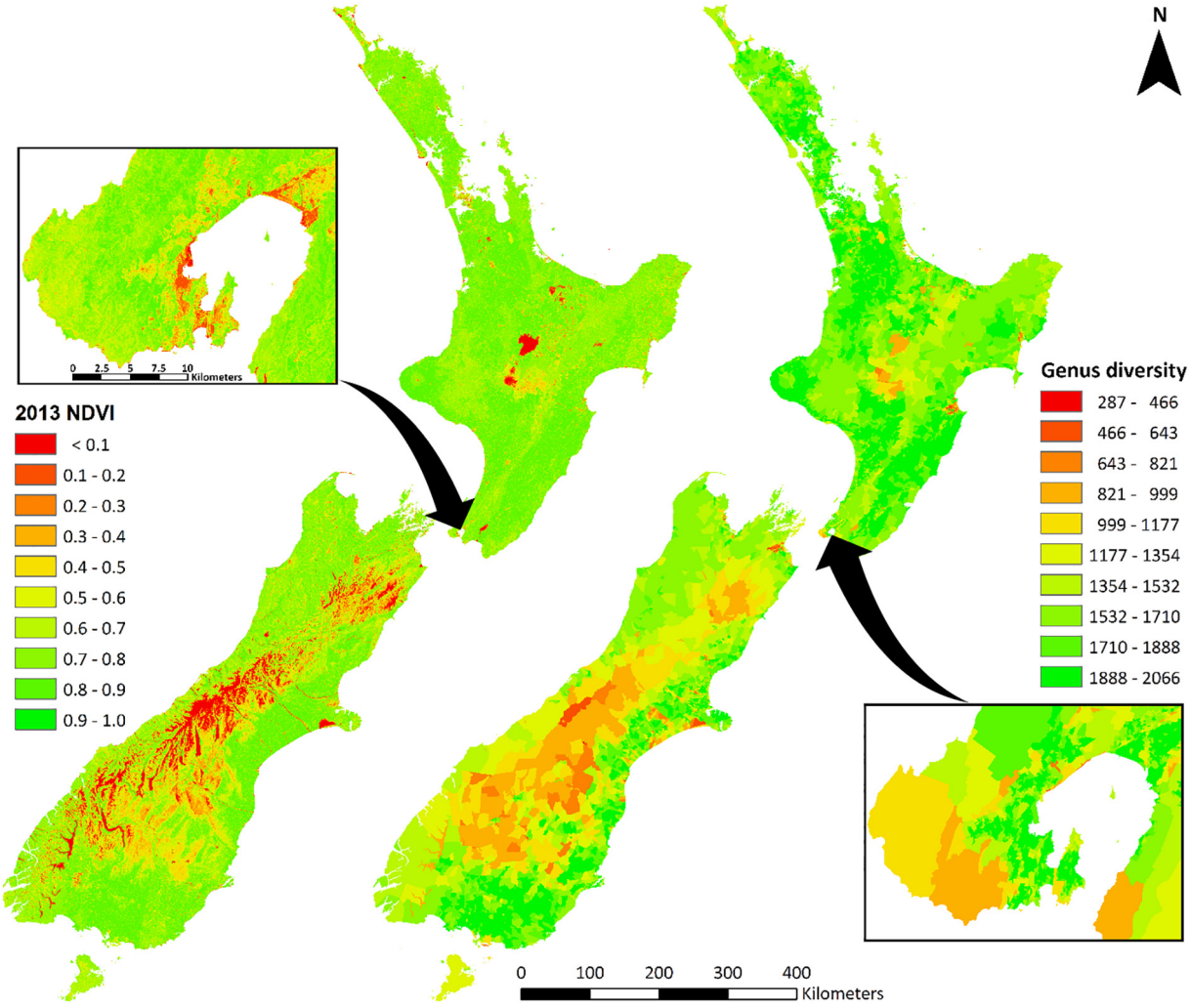


Fig. 2. Normalized difference vegetation index (NDVI) and the number of plant genera at the meshblock level. Insert shows detail for Wellington, New Zealand's capital.

variables(Salmond and Crampton 2012) that ranks meshblocks into deciles from 1 (least deprived) to 10 (most deprived). Finally, we estimated meshblock-level road density and mean annual nitrogen dioxide concentrations. The nitrogen-dioxide estimates were developed by the National Institute of Water and Atmospheric Research and are based on the sum of two semi-empirical models: an urban background estimate and a local increment from nearby roads. Nitrogen-dioxide measurements were taken between 2006 and 2016 (data available from the authors).

2.5. Statistical analysis

We estimated logit models of ALL incidence including cases of ALL that occurred from age two to age five (210 cases) and exposure metrics calculated from birth to age two. This approach ensured that exposure metrics only reflected pre-diagnostic exposure. Excluding post-diagnostic exposure was especially important in this study, because in New Zealand pediatric cancer treatment is available at three main hospitals, so families often move for treatment.

We evaluated three types of exposures (mean, minimum, and maximum) at four taxonomic scales (species, genus, family, and order). Model selection involved backwards selection with progressively lower p-value thresholds (final threshold: $p < 0{\cdot}1$, two sided test). We systematically reintroduced insignificant covariates (particularly those that were associated with ALL in past studies) and retained them, if coefficients of interest changed >10%. To gain further insight into any dose-response relationships, we re-estimated our models with biodiversity represented as tertiles rather than continuously (The cutoffs for categorical exposure metrics and covariates, as well as the number of cases and non-cases in each category, are shown in Table 1). Finally, we re-estimated continuous and categorical models using pre-natal exposure metrics.

As a validity check, we re-estimated models with all non-ALL cancers as the case group (we would not expect exposure to plant diversity to protect against non-ALL cancers). In addition, for ease of interpretation, and to allow us to compare plant diversity metrics at different taxonomic scales, we standardized plant-diversity and greenness metrics by subtracting the mean and dividing by the standard deviation. Therefore, odds ratios for plant-diversity metrics (Tables 3 and 4) denote the effect of a 1-SD increase in exposure on the probability of developing ALL. Finally, to ensure that results were not an artifact of our choice of exposure window, we re-estimated models with three other exposure windows: 0–1 years, 0–3 years, and 0–4 years. As the exposure window lengthened, we dropped a correspondingly higher number of cases of ALL to prevent the inclusion of post-diagnostic exposure in our models. For example, in models that used an exposure window from birth to a child's fourth birthday, we only included the 60 cases of ALL that occurred from age 4–5.

3. Results

Table 2 shows descriptive statistics for select exposures and covariates. Children with ALL were more likely to have European ancestry,

Table 1
Cutoffs and number of cases and non-cases for categorical plant-diversity metrics and covariates. Following IDI protocols, all sample sizes have been rounded to the nearest multiple of three.

Variable	33rd percentile	67th percentile	# of cases	# of non-cases
REFERENCE: No older siblings			147	461,691
1 older sibling			84	262,017
>1 older sibling			30	166,116
REFERENCE: Mother's age T1	27	33	81	325,350
Mother's age T2			99	312,912
Mother's age T3			84	251,559
REFERENCE: Sometimes rural resident			24	103,890
Always urban resident			237	785,931
REFERENCE: Max # genera age 0 to 2 T1	2013	2066	72	219,573
Max # genera age 0 to 2 T2			81	244,908
Max # genera age 0 to 2 T3			96	425,343

live in an urban area, have fewer older siblings, and were exposed to fewer plant genera than children without ALL.

Exposure metrics based on the maximum number of plant genera were most protective of ALL (Fig. 3). Therefore, we used genus-level metrics in all subsequent models. Exposure metrics based on mean and minimum exposure to plant diversity were also significantly associated with a reduction in ALL risk (Table 3). However, the overall explanatory power of the mean and minimum models were lower than equivalent maximum-exposure models, and the statistical significance and magnitude of mean and minimum exposures were also lower. When split into tertiles, exposure to maximum number of plant genera was associated with ALL in a dose-dependent manner (Table 4).

A number of insignificant results are notable. Overall greenness (NDVI), social deprivation (NZDep), road density, nitrogen dioxide, population density, and ethnicity, were not significantly associated with ALL, nor did inclusion of any of these insignificant variables induce a greater than 10% change in any coefficients of interest. In addition, plant diversity was not significantly associated with other (non-ALL) cancers (Supplementary Table S1), nor was prenatal exposure significantly associated with ALL (Supplementary Table S2).

Plant diversity remained protective of ALL, when we used different exposure windows (Supplementary Table S3), although there was some expected loss of significance as the exposure window increased, the number of cases of ALL declined, and the power of the analysis was reduced. In the model that only considered exposure during the first year of life (exposure 0–1 in Supplementary Table S3), the protective effect of plant diversity also declined, in both magnitude and significance, despite the inclusion of 39 more cases than our base model (Table 3).

Table 2
Descriptive statistics for children born in New Zealand from 1998 to 2013 separated by children who developed ALL ($n = 264$)[a] and those that did not ($n = 898{,}762$).[a]

Variable	*CASES ($n = 264$)*[a]		NON-CASES ($n = 898{,}762$)	
	Mean	SD	Mean	SD
Ethnicity: European	65·1%	–	60·4%	–
Ethnicity: Māori	17·4%	–	20·2%	–
Ethnicity: Pacific	5·30%	–	9·04%	–
Ethnicity: Asian	9·85%	–	7·99%	–
Ethnicity: MELAA	1·14%	–	0·85%	–
Ethnicity: Other	0·379%	–	0·604%	–
Always lived in urban area	89·8%	–	83·3%	–
Always lived in rural area	8·70%	–	8·91%	–
Number of older siblings	0·636	0.921	0·791	1·09
Mother's age	30.3	5.80	29·6	6·13
Maximum # plant genera	1946	219	1991	167
Mean # plant genera	1930	220	1951	183
Minimum # plant genera	1914	233	1901	235
Maximum NDVI	0.529	0.119	0·547	0·119
NZDep	5·60	2.90	6·01	2·78

[a] Following IDI protocols, all sample sizes have been rounded to the nearest multiple of three.

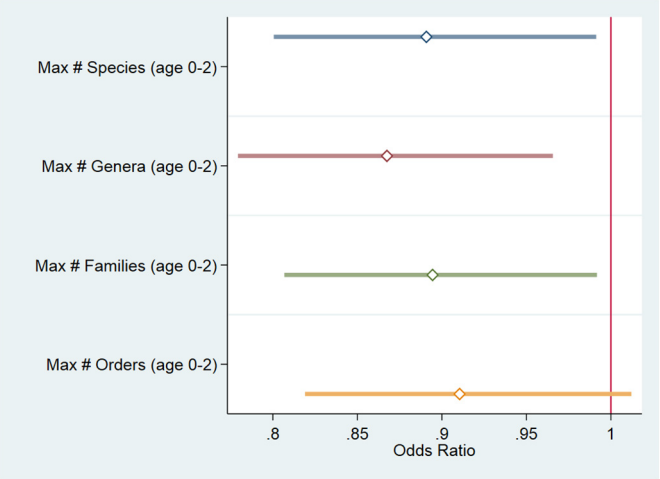


Fig. 3. The association between exposure to plant diversity (age 0–2) at four taxonomic and ALL among children born in New Zealand 1998–2013 ($n = 899{,}126$, 210 cases of ALL).

4. Discussion

This large New Zealand birth-cohort study is the first to show that plant diversity is associated, in a dose-dependent way, with a reduced incidence of ALL. The protective effect was relatively large, with the highest tertile of plant diversity associated with a reduction in risk of 35% (95% CI: 11–55%). For comparison, having more than one older sibling was associated with a 47% reduction in risk, and having a mother in the top age tertile (>33 years old) was associated with a 48% increase in risk. No association with plant biodiversity was observed for all other cancers combined suggesting that these effects are specific to ALL. In addition, prenatal exposure was not associated with ALL incidence, which suggests that *in utero* exposure to plant diversity may be less important. Consistent with other studies, we also found that having an older mother (Hemminki et al. 1999) and living in an urban area (Stiller et al. 2008) were risk factors for ALL, whilst older siblings were protective (Ajrouche et al. 2015). If the associations reported here are causal then this is an important finding, as there are currently no public-health interventions to reduce ALL risk, and plants are a modifiable component of the environment.

Plant diversity, which is known to be associated with diverse environmental microbiota (Kowalchuk et al., 2002; Lymperopoulou et al., 2016), may protect against ALL by promoting immune maturation through changes in the human microbiota (as suggested by the biodiversity hypothesis). This mechanistic pathway is supported by a recent case-control study (Rajagopala et al. 2016), which found marked pre-treatment differences in the gut microbiota of children with ALL. However, it is possible that this is caused by the disease process rather than it being involved in the causal pathway of ALL. Our findings are also consistent with a previous study conducted by this research team, which found that children who lived in more biodiverse neighborhoods were less likely to develop childhood asthma—another condition with an immune component that is influenced by the gut microbiota (Donovan et al. 2018).

Exposure to plants can have other beneficial effects including stress reduction, improved air quality, and better social connections. These mechanisms likely play a role in the protective effect of the natural environment on other health outcomes, and we cannot dismiss the possibility that exposure to plants, rather than their microbial communities, are also directly protective of ALL. However, greenness (NDVI) was not associated with a reduction in ALL risk, which suggest that plant

Table 3
Association between exposure to maximum, mean, and minimum plant diversity and ALL incidence among children born New Zealand 1998–2013 ($n = 899,126$). Following IDI protocols, all sample sizes have been rounded to the nearest multiple of three. OR = Odds Ratio; CI=Confidence Interval.

Variables	OR (95% CI)	p-Value	OR (95% CI)	p-Value	OR (95% CI)	p-Value
REFERENCE: No older siblings	1.00		1.00		1.00	
1 older sibling	0.921 (0.696–1.218)	0.564	0.927 (0.705–1.218)	0.587	0.927 (0.705–1.218)	0.587
>1 older sibling	0.497 (0.329–0.752)	<0.001	0.529 (0.356–0.785)	0.0016	0.529 (0.356–0.785)	0.0016
REFERENCE: Mother's age T1	1.00		1.00		1.00	
Mother's age T2	1.382 (1.017–1.879)	0.039	1.369 (1.016–1.844)	0.039	1.369 (1.016–1.844)	0.039
Mother's age T3	1.52 (1.098–2.104)	0.012	1.500 (1.094–2.057)	0.012	1.5 (1.094–2.057)	0.012
REFERENCE: Sometimes rural resident	1.00		1.00		1.00	
Always urban resident	1.495 (0.959–2.33)	0.076	1.447 (0.937–2.235)	0.095	1.447 (0.937–2.235)	0.095
Max # genera age 0 to 2	0.867 (0.779–0.966)	0.009				
Mean # genera age 0 to 2			0.877 (0.785–0.979)	0.019		
Min # genera age 0 to 2					0.893 (0.797–1.001)	0.052
R-squared	0.0055		0.0053		0.0053	

diversity is not merely a proxy for exposure to the natural environment. The insignificance of greenness (NDVI) suggests that public-health interventions to reduce ALL risk through plant exposure, should prioritize plant diversity over overall greenness. Finally, neither nitrogen dioxide nor road density (both proxies for traffic-related air pollution) were significantly associated with ALL, nor did their inclusion in regression models change coefficients of interest by >10%, which suggests that the association between plant diversity and ALL is not confounded by improved air quality.

Our finding that maximum (rather than mean or minimum) exposure to plant diversity offered the greatest protection suggests that public-health interventions may be most effective, if they provide exposure to high plant diversity, even for shorter periods of time, rather than sustained exposure to lower-level diversity. However, our results do not provide any guidance on the optimal duration or intensity of this exposure. When split into tertiles, plant diversity was only protective in the top tertile, which suggests that plant diversity may only protect against ALL above some minimum threshold.

We also found that diversity at the genus level was most protective of ALL and explained more of the variance in ALL incidence than equivalent metrics based on plant species, family, or order. The poorer performance of metrics at the family and order level is expected, because information is lost as the taxonomic scale of metrics coarsens. The poorer performance of species-level metrics was somewhat unexpected and suggests that differences in microbial communities among species of the same genus may be insufficient to provoke unique immune responses. However, it is not clear what elements of this genus-level diversity offer the most protection. For example, is structural plant diversity most important, or is diversity in the chemical composition of plant leaves more important?

When we shortened the exposure window to just the first year of life, plant diversity remained protective of ALL, but the magnitude and significance of the association declined, which suggests that exposure beyond the first year of life is protective. When we lengthened the exposure window beyond the first two years of life, the association between plant diversity and ALL declined in significance, although this is likely due, at least in part, to a reduction in statistical power.

Our study has several limitations. This was an observational study, so we were unable to prove a causal relationship between exposure to plant diversity and ALL risk. The observed associations should therefore be interpreted with care until future experimental studies confirm a causal link between plant diversity and ALL. Our exposure metrics were based on residential history, which is at the meshblock level rather than precise address. In consequence, all our exposure metrics are subject to measurement error. This is especially true for children living in large rural meshblocks. In addition, the accuracy of residential data within the IDI depend on how often an individual has contact with bureaucratic systems. These contacts are likely to be frequent around a child's birth, but may be less so during the early pre-natal period and as a child ages. Therefore, exposure misclassification may be greater for prenatal exposure and exposures later in a child's life. Our plant-diversity metric is based upon national totals in each land-cover type, but it is unlikely that all plant species found in a specific land-cover type are also found in a given meshblock. This is a significant limitation. Future studies could valuably focus on improving biodiversity-exposure metrics. In addition, plant-diversity data come from multiple surveys, so it is possible that some areas are over sampled. However, most of these limitations would likely result in random, as opposed to systematic, measurement error and would, therefore, result in less precise, as opposed to biased, effect estimates.

In conclusion, this study suggests that exposure to plant biodiversity may be protective of childhood ALL. If true, the loss of global biodiversity may, at least in part, explain the increase in childhood ALL incidence in many parts of the world. Public-health interventions to increase children's exposures to plant diversity may be a feasible option to reduce childhood ALL risk. Finally, our findings provide support for the

Table 4
Association between maximum exposure to plant diversity (represented continuously and in tertiles) and ALL incidence among children born New Zealand 1998–2013 (n = 899,126). Following IDI protocols, all sample sizes have been rounded to the nearest multiple of three. OR = Odds Ratio; CI=Confidence Interval.

Variables	OR (95% CI)	p-Value	OR (95% CI)	p-Value
REFERENCE: No older siblings	1.00		1	
1 older sibling	0.921 (0.696–1.218)	0.564	0.923 (0.702–1.213)	0.566
>1 older sibling	0.497 (0.329–0.752)	<0.001	0.525 (0.354–0.779)	0.0014
REFERENCE: Mother's age T1	1.00		1	
Mother's age T2	1.382 (1.017–1.879)	0.039	1.373 (1.019–1.85)	0.037
Mother's age T3	1.52 (1.098–2.104)	0.012	1.511 (1.102–2.072)	0.011
REFERENCE: Sometimes rural resident	1.00		1	
Always urban resident	1.495 (0.959–2.33)	0.076	1.525 (0.987–2.355)	0.057
Max # genera age 0 to 2	0.867 (0.779–0.966)	0.009		
REFERENCE: Max # genera age 0 to 2 T1			1	
Max # genera age 0 to 2 T2			0.982 (0.712–1.354)	0.911
Max # genera age 0 to 2 T3			0.645 (0.47–0.885)	0.0066
R-squared	0.0055		0.0065	

broader biodiversity hypothesis, which posits that global loss of biodiversity is, at least in part, responsible for an increase in the incidence of immune diseases, especially in high-income countries.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2020.144627.

CRediT authorship contribution statement

GHD developed the research idea with JD, conducted the analysis and took the lead on writing the paper. DG led the geo-spatial analysis and wrote the geo-spatial section of the methods. AM consulted on the statistical analysis and edited several drafts of the paper. RW consulted on clinical issues and edited several drafts of the paper. *CF* contributed to the geo-spatial analysis. JD helped develop the research question, wrote sections of the paper, and edited several drafts.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was partially supported by the Organization for Economic Cooperation and Development's Cooperative Research Program.

The results in this paper are not official statistics. They have been created for research purposes from the Integrated Data Infrastructure (IDI), managed by Statistics New Zealand. The opinions, findings, recommendations, and conclusions expressed in this paper are those of the authors, not Statistics NZ. Access to the anonymized data used in this study was provided by Statistics NZ under the security and confidentiality provisions of the Statistics Act 1975. Only people authorized by the Statistics Act 1975 are allowed to see data about a particular person, household, business, or organization, and the results in this paper have been confidentialized to protect these groups from identification and to keep their data safe. Careful consideration has been given to the privacy, security, and confidentiality issues associated with using administrative and survey data in the IDI. Further detail can be found in the Privacy impact assessment for the Integrated Data Infrastructure available from www.stats.govt.nz.

References

Ajrouche, R., Rudant, J., Orsi, L., Petit, A., Baruchel, A., Lambilliotte, A., et al., 2015. Childhood acute lymphoblastic leukaemia and indicators of early immune stimulation: the Estelle study (SFCE). Br. J. Cancer 112, 1017.

Barrington-Trimis, J.L., Cockburn, M., Metayer, C., Gauderman, W.J., Wiemels, J., McKean-Cowdin, R., 2017. Trends in childhood leukemia incidence over two decades from 1992 to 2013. Int. J. Cancer 140, 1000–1008.

Carney, K.M., Matson, P.A., 2006. The influence of tropical plant diversity and composition on soil microbial communities. Microb. Ecol. 52, 226–238.

Chander, G., Markham, B.L., Helder, D.L., 2009. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. Remote Sens. Environ. 113, 893–903.

Cho, I., Blaser, M., 2012. The human microbiome: at the interface of health and disease. Nat. Rev. Genet. 13, 260.

Dallimer, M., Irvine, K.N., Skinner, A.M., Davies, Z.G., Rouquette, J.R., Maltby, L.L., et al., 2012. Biodiversity and the feel-good factor: understanding associations between self-reported human well-being and species richness. BioScience 62, 47–55.

Donovan, G.H., Gatziolis, D., Longley, I., Douwes, J., 2018. Vegetation diversity protects against childhood asthma: results from a large New Zealand birth cohort. Nature Plants 4, 358–364.

Gollwitzer, E.S., Marsland, B.J., 2015. Impact of early-life exposures on immune maturation and susceptibility to disease. Trends Immunol. 36, 684–696.

Greaves, M., 2018. A causal mechanism for childhood acute lymphoblastic leukaemia. Nat. Rev. Cancer 18, 471–484.

Haahtela, T., Holgate, S., Pawankar, R., Akdis, C.A., Benjaponpitak, S., Caraballo, L., et al., 2013. The biodiversity hypothesis and allergic disease: world allergy organization position statement. World Allergy Organization Journal 6, 1.

Hanski, I., von Hertzen, L., Fyhrquist, N., Koskinen, K., Torppa, K., Laatikainen, T., et al., 2012. Environmental biodiversity, human microbiota, and allergy are interrelated. Proc. Natl. Acad. Sci. 109, 8334–8339.

He, Y., Wu, W., Zheng, H.M., Li, P., McDonald, D., Sheng, H.F., et al., 2018. Regional variation limits applications of healthy gut microbiome reference ranges and disease models. Nat. Med. 24, 1532–1535.

Hemminki, K., Kyyrönen, P., Vaittinen, P., 1999. Parental age as a risk factor of childhood leukemia and brain cancer in offspring. Epidemiology 10, 271–275.

Hjalmars, U., Gustafsson, G., 1999. Higher risk for acute childhood lymphoblastic leukaemia in Swedish population centres 1973-94. Br. J. Cancer 79, 30–33.

Inaba, H., Greaves, M., Mullighan, C.G., 2013. Acute lymphoblastic leukaemia. Lancet 381, 1943–1955.

Isolauri, E., 2001. Probiotics in human disease. Am. J. Clin. Nutr. 73, 1142S–1146S.

Jenmalm, M.C., 2011. Childhood immune maturation and allergy development: regulation by maternal immunity and microbial exposure. Am. J. Reprod. Immunol. 66, 75–80.

Kowalchuk GA, Buma DS, de Boer W, Klinkhamer PG, van Veen JA. Effects of aboveground plant species composition and diversity on the diversity of soil-borne microorganisms. Antonie Van Leeuwenhoek 2002; 81: 509.

Land Information New Zealand, 2012. Land Cover Database Version 4.1. vol. 2019.

Liddicoat, C., Bi, P., Waycott, M., Glover, J., Lowe, A.J., Weinstein, P., 2018. Landscape biodiversity correlates with respiratory health in Australia. J. Environ. Manag. 206, 113–122.

Liddicoat, C., Sydnor, H., Cando-Dumancela, C., Dresken, R., Liu, J., Gellie, N.J.C., et al., 2020. Naturally-diverse airborne environmental microbial exposures modulate the gut microbiome and may provide anxiolytic benefits in mice. Sci. Total Environ. 701, 134684.

Linabery, A.M., Ross, J.A., 2008. Trends in childhood cancer incidence in the U.S. (1992–2004). Cancer 112, 416–432.

Lymperopoulou, D.S., Adams, R.I., Lindow, S.E., 2016. Contribution of vegetation to the microbial composition of nearby outdoor air. Applied environmental microbiology 81, 509.

Milne, B.J., Atkinson, J., Blakely, T., Day, H., Douwes, J., Gibb, S., et al., 2019. Data resource profile: the New Zealand integrated data infrastructure (IDI). Int. J. Epidemiol. 48 (3), 677-677E.

Mori, H., Colman, S.M., Xiao, Z., Ford, A.M., Healy, L.E., Donaldson, C., et al., 2002. Chromosome translocations and covert leukemic clones are generated during normal fetal development. Proc. Natl. Acad. Sci. 99, 8242–8247.

Phillips, J.N., Berlow, M., Derryberry, E.P., 2018. The Effects of Landscape Urbanization on the Gut Microbiome: An Exploration Into the Gut of Urban and Rural White-crowned Sparrows. vol. 6.

Rajagopala, S.V., Yooseph, S., Harkins, D.M., Moncera, K.J., Zabokrtsky, K.B., Torralba, M.G., et al., 2016. Gastrointestinal microbial populations can distinguish pediatric and adolescent acute lymphoblastic leukemia (ALL) at the time of disease diagnosis. BMC Genomics 17, 635.

Salmond, C.E., Crampton, P., 2012. Development of New Zealand's deprivation index (NZDep) and its uptake as a national policy tool. Canadian Journal of Public Health 103, S7–S11.

Schmitt, S., Pouteau, R., Justeau, D., de Boissieu, F., Birnbaum, P., Golding, N., 2017. SSDM: an R package to predict distribution of species richness and composition based on stacked species distribution models. Methods Ecol. Evol. 8, 1795–1803.

Stiller, C.A., Kroll, M.E., Boyle, P.J., Feng, Z., 2008. Population mixing, socioeconomic status and incidence of childhood acute lymphoblastic leukaemia in England and Wales: analysis by census ward. Br. J. Cancer 98, 1006–1011.

Vorholt, J.A., 2012. Microbial life in the phyllosphere. Nat Rev Microbiol 10, 828–840.

Wang, R., Wiemels, J.L., Metayer, C., Morimoto, L., Francis, S.S., Kadan-Lottick, N., et al., 2017. Cesarean section and risk of childhood acute lymphoblastic leukemia in a population-based, record-linkage study in California. Am. J. Epidemiol. 185, 96–105.