Infection, Genetics and Evolution 61 (2018) 189–196

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Research paper

The macroecology of cancer incidences in humans is associated with large-scale assemblages of endemic infections

Camille Jacqueline[a,b,*], Jessica L. Abbate[b,c], Gabriele Sorci[d], Jean-François Guégan[b], Fréderic Thomas[a,b], Benjamin Roche[a,c,e]

[a] CREEC, 911 Avenue Agropolis, BP 64501, 34394 Montpellier, Cedex 5, France
[b] MIVEGEC, UMR IRD/CNRS/Université de Montpellier, 911 Avenue Agropolis, BP 64501, 34394 Montpellier Cedex 5, France
[c] International Center for Mathematical and Computational Modeling of Complex Systems (UMI IRD/Sorbonne Université UMMISCO), 32 Avenue Henri Varagnat, 93143 Bondy Cedex, France
[d] BiogéoSciences, CNRS UMR 6282, Université de Bourgogne, 6 Boulevard Gabriel, 21000 Dijon, France
[e] Departamento de Etología, Fauna Silvestre y Animales de Laboratorio, Facultad de Medicina Veterinaria y Zootecnia, Universidad Nacional Autónoma de México (UNAM), Ciudad de México, México.

ARTICLE INFO

Keywords:
Neglected diseases
Biomes
Human cancer incidences
Data mining
Public health strategies
Pathogen-cancer interactions

ABSTRACT

It is now well supported that 20% of human cancers have an infectious causation (i.e., oncogenic agents). Accumulating evidence suggests that aside from this direct role, other infectious agents may also indirectly affect cancer epidemiology through interactions with the oncogenic agents within the wider infection community. Here, we address this hypothesis via analysis of large-scale global data to identify associations between human cancer incidence and assemblages of neglected infectious agents. We focus on a gradient of three widely-distributed cancers with an infectious cause: bladder (~2% of recorded cancer cases are due to *Shistosoma haematobium*), liver (~60% consecutive to Hepatitis B and C infection) and stomach (*Helicobacter pylori* is associated with ~70% of cases). We analyzed countries in tropical and temperate regions separately, and controlled for many confounding social and economic variables. First, we found that particular assemblages of bacteria are associated with bladder cancer incidences. Second, we observed a specific and robust association between helminths and liver cancer incidences in both biomes. Third, we show that certain assemblages of viruses may facilitate stomach cancer in tropical area, while others protect against its development in temperate countries. Finally, we discuss the implications of our results in terms of cancer prevention and highlight the necessity to consider neglected diseases, especially in tropics, to adapt public health strategies against infectious diseases and cancer.

1. Introduction

While cancer remains one of the main causes of death in Western countries (Ferlay et al., 2010), its burden is increasing in low- and middle-income countries (Magrath et al., 2013). Although treatments, including chemo-, radio- and/or immunotherapy, have resulted in a slight decline in death rates worldwide, cancer incidence rates have remained stable over the past 10 years (Siegel et al., 2013). In this context, prevention seems currently the most efficient strategy to reduce the impact of cancer on populations. The term "infectious agents" describes the transmissible organisms that require living in or on another organism (the "host") to complete its lifecycle – a process which decreases host fitness – and includes virus, bacteria, fungi, protozoans, helminths, etc… Since the 20th century, numerous infectious agents have been recognized as risk factors for the development of several cancers (i.e., acting through inflammation or introduction of foreign DNA into cells; henceforth referred to "oncogenic agents") (Zur Hausen and De Villiers, 2015). For example, current evidence links Epstein–Barr virus (EBV), Hepatitis B and C viruses (HBV, HCV), the bacteria *Helicobacter pylori*, human papillomavirus (HPV), and the trematode *Schistosoma haematobium* to cancers of the lymph nodes, liver, stomach, cervix and bladder respectively. Consequently, vaccination against oncogenic agents is currently being considered as a potential key tool to prevent these cancers, such as the HPV vaccine which offers a relative protection against cervical cancer (Paavonen et al., 2009).

Accumulating evidence suggests that aside from this direct role, infectious agents could also be indirectly involved in cancer epidemiology through interactions with oncogenic agents in infra-

* Corresponding author at: 911 avenue Agropolis, 34000 Montpellier, France.
E-mail address: camille.jacqueline@ird.fr (C. Jacqueline).

https://doi.org/10.1016/j.meegid.2018.03.016
Received 4 November 2017; Received in revised form 19 March 2018; Accepted 21 March 2018
Available online 23 March 2018

communities (reviewed in Jacqueline et al. (2017)). Indeed, hosts are commonly infected by multiple infectious species simultaneously (Read and Taylor, 2001), and within-host competition for resources and through immunity has been fairly well-established (reviewed in Mideo (2009)). A number of infectious organisms are known to impair immune system homeostasis with both positive and negative consequences for their competitors, especially through the well-known trade-off between the Th1/Th17 and Th2 immune pathways (Pedersen and Fenton, 2007). For example, helminth infections are often responsible for an up-regulation of the Th2 response, which has been linked to tuberculosis reactivation and increase in likelihood of HIV infection (Borkow et al., 2001). Furthermore, these interactions at individual scale are susceptible to have consequences on persistence and transmission dynamics of the infectious agents at the population level (Ezenwa and Jolles, 2014).

A few studies have explicitly considered oncogenic agents as part of an infectious community. The majority of the literature available on interactions between oncogenic and non-oncogenic agents concerns species that are widely distributed and with high prevalence. For instance, *Plasmodium falciparum*, the agent of malaria, increases the replication of EBV, an oncogenic virus associated with Burkitt Lymphoma, through its impairment of an effective immune response (Chêne et al., 2007; Morrow et al., 1976). Epidemiological studies have also shown that infection with *Chlamydia trachomatis* increases the risk for persistence of HPV infection, some lineages of which can lead to cervical cancer (Silins et al., 2005). However, these non-oncogenic agents, with high prevalence, may not be representative of the whole diversity of infectious organisms, especially in tropical countries that are affected by a high number of endemic and rare diseases that are relatively under-studied (Hotez et al., 2007).

Here, we explore the hypothesis that co-infections with endemic infections, in both tropical and temperate latitudes, modify the persistence of oncogenic agents and thus interact with the development of some infectious cancer in human population. For most countries, prevalence of oncogenic agents is not available for the whole population and thus we postulate that infectious agents favoring persistence of oncogenic agents should co-localize in countries with higher cancer incidences. We focus on three widely-distributed cancers (bladder, liver and stomach), for which an infectious causation is widely recognized and for which sufficient large-scale data are available. These three examples allow considering oncogenic agents that belong to different taxonomic guilds (helminthes, viruses and bacteria) as well as a gradient in the frequency with which each cancer is known to have an infectious origin (from 2% of recorded cases for bladder cancer, to 60% for liver cancer and 70% for stomach cancer). By compiling and analyzing a global database, we identify associations between the country-level incidence of cancer and the presence or absence of infectious agents in each country. We then present hypotheses regarding the mechanism behind each association, and discuss the potential consequences of our findings in terms of cancer prevention strategies.

2. Material & methods

2.1. Data overview and inclusion criteria

Our cancer data were obtained from the International Agency for Research on Cancer (IARC GLOBOCAN project, 2012, http://globocan.iarc.fr/). Among 48 human cancers in women and men across 184 countries, we selected three cancers of interest (bladder, liver and stomach) because they are recognized to have an infectious causation. First, the bacterium *H. pylori* is responsible for 80% of stomach adenocarcinomas (Zur Hausen, 2009), which represent 90% of stomach cancers (Brenner et al., 2009). Second, 75% of liver cancers are hepatocellular carcinomas (HCCs) (Parkin, 2001) which have been linked in more than 80% of cases to HCV and HBV infection (Parkin, 2006). Finally, squamous cell carcinomas (SCCs), representing 5% of bladder cancers (Kantor et al., 1988), have been associated with the helminth *Schistosoma haematobium* in approximately 30% of cases (Mostafa and Sheweita, 1999). In our study, we used age-standardized incidence (ASI), which represents the raw cancer incidence when country age structure is extrapolated to a standard one (World Standard Population (Doll et al., 1996)).

Our analyses on infectious species relied on the GIDEON database (Global Infectious Diseases and Epidemiology Network, http://web.gideononline.com/). This database contained a presence/absence matrix for a total of 370 human infectious agents across 224 countries (2004 update). We removed all infectious agents that were present or absent in all countries, because they did not add any discriminating information, as well as infections caused by fungi, protozoans and arthropods guilds which were represented by less than 10 infectious agents respectively. This yielded a data subset with infectious agents belonging to three main guilds (viruses, bacteria and helminths). It included very rare diseases such as Buruli ulcer (3000 cases/y worldwide; www.who.int/gho/neglected_diseases/buruli_ulcer) and tick-borne encephalitis (11,000 cases/y worldwide; www.who.int/immunization/topics/tick_encephalitis) to more widespread diseases affecting a higher number of persons such as leshmaniasis and echinococcosis (1 million cases/y worldwide; http://www.who.int/echinococcosis/).

Finally, we constituted a database of 24 potential confounding variables for 167 countries from the Food and Agriculture Organization (FAO), World Health Organization (WHO) and the World Bank (WB) to consider the diversity of social factors and economic levels (Supplementary Table S1 for a description).

The dataset was analyzed separately between tropical (geographically situated between the two tropics) and temperate (above or below tropics) countries in order to better control for the huge environmental (in addition to social and economical) disparities between these two biomes (distinct biological communities formed in response to a shared physical climate (Cain et al., 2011)). In fact, tropical countries have a higher abundance/richness of infectious species (Guernier et al., 2004) and are associated on average with lower socio-economic wealth (World Bank 2016) compared to temperate countries. We confirmed that these two groups of countries, made on geographical assumptions, correspond to two distinct biomes which present distinct patterns of infectious species richness (Supplementary Fig. S1) and composition (Fig. 1A) as well as socio-economical disparities (Fig. 1B).

2.2. Assessment of infectious agent assemblages

Our database of infectious agents, after removing non-relevant species as previously described, contained presence/absence values for 101 human infectious agents across 103 tropical countries and 88 infectious agents across 74 temperate countries. We first aimed to characterize infectious assemblages within each guild (i.e., viruses, bacteria and helminths). To do so, we ran Multiple Correspondence Analysis (MCA function; package *FactomineR*, Lê et al., 2008) in order to reduce the number of variables to a set of dimensions representing the presence/absence data of all infectious agents within each guild. The composition of these dimensions was then examined as a synthetic measure of assemblage structure. The number of dimensions considered was calculated according to the Kaiser's criterion (dimensions for which the percentage of variance explained was superior to a threshold calculated as follows: 100%/total number of dimensions estimated by MCA analysis (Kaiser, 1958)). For each of these selected dimensions, we calculated standard coordinates for each country by dividing the principal coordinates (i.e. loadings) by the square root of the dimension's eigenvalue (Husson et al., 2017). This standardization allowed considering each dimension as independent variables.

We followed a similar methodology for the 24 confounding variables described previously (substituting the Multiple Correspondence Analysis with a Principal Component Analysis, as our confounding