

BRIEF REPORT

Open Access



Power-law scaling in intratumoral microbiota of colorectal cancer

Nikolas Dovrolis^{1*}, Maria Gazouli^{1*}, François Rigal^{2,3}, Robert J. Whittaker^{4,5}, Thomas J. Matthews^{3,6}, Konstantinos Georgiou⁷, George Theodoropoulos⁷ and Kostas A. Triantis^{8*}

Abstract

It has recently been proposed that the study of microbial dynamics in humans may gain insights from island biogeographical theory. Here, we test whether the diversity of the intratumoral microbiota of colorectal cancer tumors (CRC) follows a power law with tumor size akin to the island species-area relationship. We confirm a direct correlation between the quantity of Amplicon Sequence Variants (ASVs) within CRC tumors and tumor sizes, following a (log)power model, explaining 47% of the variation. Understanding the processes involved, potentially through the analogy of tumors and islands, may ultimately contribute to future clinical and therapeutic strategies.

Keywords Intratumoral microbiome, Cancer, Island species–area relationship

*Correspondence:

Nikolas Dovrolis
ndovroli@med.uoa.gr
Maria Gazouli
mgazouli@med.uoa.gr
Kostas A. Triantis
ktriantis@biol.uoa.gr

¹Department of Basic Medical Sciences, Laboratory of Biology, Medical School, National and Kapodistrian University of Athens, Athens 11527, Greece

²Institut Des Sciences Analytiques et de Physico Chimie pour L'environnement et les Matériaux, CNRS - Université de Pau et des Pays de l'Adour - E2S UPPA, UMR5254, Pau 64000, France

³CE3C – Centre for Ecology, Evolution and Environmental Changes/ Azorean Biodiversity Group, Faculty of Agricultural Sciences and Environment, CHANGE – Global Change and Sustainability Institute and Universidade dos Açores, Angra do Heroísmo, Açores PT-9700-042, Portugal

⁴School of Geography and the Environment, University of Oxford, Oxford, UK

⁵Center for Macroecology, Evolution and Climate, GLOBE Institute, University of Copenhagen, Copenhagen, Denmark

⁶GEES (School of Geography, Earth and Environmental Sciences, Birmingham Institute of Forest Research, University of Birmingham, Birmingham, UK

⁷1st Department of Propaedeutic Surgery, Medical School, National Kapodistrian University of Athens, Hippocratio Hospital, Athens 11527, Greece

⁸Department of Ecology and Taxonomy, National & Kapodistrian University of Athens, Athens 11527, Greece

Power laws are found in numerous natural, social, and artificial systems, phenomena [1–4] and are particularly prevalent in biology, ecology, and biogeography, demonstrating that evolutionary, ecological, and physiological constraints apply consistently to genomes, cells, organisms, communities, and ecosystems across vast orders of scale [1, 5–8]. One of the best-established ecological laws is the island species–area relationship (ISAR), wherein the number of species ('species richness') occurring per island increases per island area. This pattern has been substantiated by studies spanning diverse spatiotemporal scales and taxa, including the microbiome [9, 10], and has been found to be best expressed as a power function which, when linearized via logarithmic transformation, is given by the equation $\log S = \log c + z \log A$, where S denotes species richness, A represents area, and c and z are fitted parameters. The slope of the ISAR, i.e. the z parameter, is often considered informative of the dominant processes of species addition [11, 12].

The seminal paper describing the Human Microbiome Project [13] references island biogeography as a driving force behind our ability to study and understand spatiotemporal microbial dynamics, while a recent study



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

proposes a positive scaling relationship between human height and gut microbiome alpha-diversity across two large, independent cohorts, controlling for a wide range of relevant covariates [14]. Notwithstanding that cancer tumor dynamics may depend on a variety of variables spanning cell proliferation inclination and genomics [15] to microbial interactions and metagenomics [16], recent research has proposed that tumors can usefully

be considered as island-like systems [17, 18]. Here, we explore this analogy by testing for an ISAR-like relationship between the microbiota, residing inside the tumor as a first step in applying island biogeography principles and theories to understanding and predicting tumor dynamics. These intratumoral microbiota may contribute to shaping the tumor microenvironment by influencing the immune environment, inflammation, and metabolic patterns of the tumor [19], hence they may provide novel ways to study cancer prognosis, development, and therapy [20].

Using a novel cohort of tumors from colorectal cancer (CRC) patients ($n=27$), a cancer type highly connected to microbial dysbiosis [21], we test for a potential relationship between the size of a tumor (measured as its diameter) and the composition of bacteria residing within the tumor. These bacteria are identified by analyzing their unique DNA sequences, called Amplicon Sequence Variants (ASVs), which act as the “species units” for this study. We hypothesize that the intratumor microbiota populations across patients will mirror the ISAR, by increasing in diversity as the tumor’s size increases, following a power law model. The power model is the most frequently preferred, both in describing ISARS and as a basis for the development of species diversity theories [11, 12].

Using a linearized power model with a logarithmic transformation of the variables (the log-log model), we identified a significant positive species–area relationship (Fig. 1A, Table S2): tumor size explaining 47% of the number of ASVs identified within CRC tumors. Rarefaction curves indicated that all tumors were sufficiently sequenced to capture total ASV within-sample richness (reached an asymptote, Figure S1). We are therefore confident that the aforementioned results are robust. However, as a sensitivity test, the analysis was rerun using rarefied species richness and the ISAR remained significant (Fig. 1B, Table S2). For all regression models, no specific deviations from residual normality and homoscedasticity and no outliers were detected (Table S2). Further sensitivity tests were run to evaluate the model’s predictive power using a k -fold cross-validation procedure. This analysis revealed a correlation between the observed and the predicted \log_{10} ASVs (Pearson’s Correlation Coefficient=0.65) with a 95% confidence interval of [0.51; 0.79] calculated across 1000 simulations (see Fig. 1C). Using rarefied richness, a mean Pearson’s Correlation Coefficient of 0.55 was obtained with a 95% confidence interval of [0.36; 0.70]. These analyses support the robustness of our findings and indicate that the relationships described here have predictive power, i.e. they may extend to other tumors.

While we caution that (i) using diameter alone to measure tumor size is a potential limitation, and (ii) varied

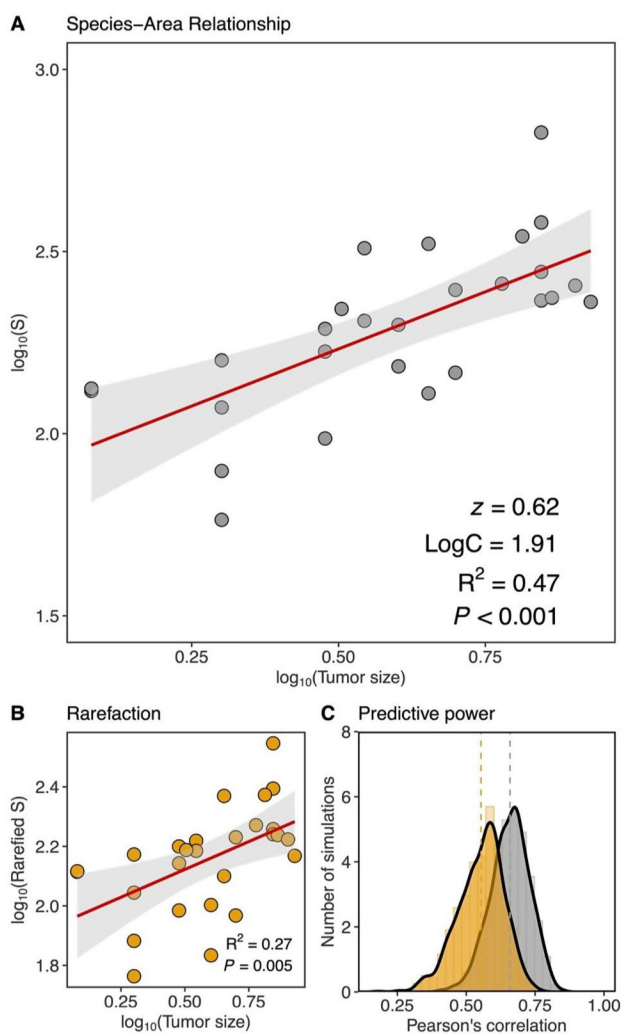


Fig. 1 (A) Relationship between the \log_{10} -species richness (Number of Amplicon Sequence Variants, ASVs) and \log_{10} -tumor size, determined using Ordinary Least Square regression; the parameter z and LogC of the model as well as the associated R^2 and the P -values of the F -statistic are given. More details on the OLS are given in Table S2. (B) Relationship between the \log_{10} of rarefied species richness (Number of Amplicon Sequence Variants, ASVs) and the \log_{10} -tumor size (R^2 and P -values of the F -statistic are given). Rarefaction was implemented with 1000 iterations. See Table S2 for details of the OLS. (C) The predictive power of the relationship in (A) and (B) evaluated using repeated k -fold cross-validation approach with 1000 simulations. The distribution of Pearson’s correlation between observed and predicted species richness values obtained for the 1000 simulations is given for both rarefied (yellow) and non-rarefied species richness (grey). Dashed lines indicate the mean correlation

individual ages and cancer stages are potential confounding factors that we did not analytically account for, the strength of the observed pattern across diverse tumors suggests further exploration of the analogy with island dynamics to hold potential.

The similarity in the value of the slope of the diversity–tumor size relationship with patterns found in oceanic islands/archipelagos [11, 12] is intriguing. While we cannot assert that tumor ontogeny resembles that of oceanic islands, our findings serve as a starting point for further investigations into the commonalities between tumor growth and the colonization and diversification dynamics observed in island ecosystems. Studies have shown that the origins of intratumoral microbiota may stem from various pathways like proximity to the mucosal barrier, colonization from neighboring normal tissues enabled by the immunosuppressive and hypoxic conditions, or hematogenous dissemination where microorganisms from other sites invade the tumor via compromised blood vessels [20, 22]. Parallels can also be drawn to the island theory in terms of the effects environmental heterogeneity has on species richness, with the more diverse islands, usually the larger ones, hosting higher richness [11, 12]. Similarly, recent studies have found that intratumoral microbiota are not randomly distributed but rather highly organized into micro-niches, with functions related to immune and epithelial cells, thereby contributing to the progression of cancer [22], since larger tumors, often characterized by irregular morphology, potentially harbor more of these niches [23]. Thus, tumors' morphology can be a potential explanation for the pattern reported herein.

In conclusion, our findings not only contribute to our understanding of tumor biology but also concur with the plea for interdisciplinary exploration by drawing parallels between tumor growth patterns and well-established ecological principles [24]. The study of the intratumoral microbiota and their dynamics holds potential to significantly impact future clinical and therapeutic endeavors [25]. Further research is warranted to unravel the mechanistic underpinnings of the observed similarities and explore the potential implications for both oncology and ecology.

Methods

Patient samples & DNA extraction

In this study, resected tumor samples from 27 CRC patients treated in the Colorectal Unit of the First Pro-paedeutic Department of Surgery, Hippocraton General Hospital, Athens, Greece were analyzed. Samples were subsequently frozen in liquid nitrogen for one hour and kept at -80°C before processing. All tumors selected ranged from 1.2 cm to 8.5 cm in diameter and were histologically inspected to validate the diagnosis. The longest

length of the tumor in the tissue removed during surgery is reported as the tumor size. DNA extraction was performed using Macherey Nucleospin Tissue (MNT, MACHEREY-NAGEL) according to the manufacturer's instructions. Sample metadata includes tumor size, patient age and sex as well as histological tumor grades, and can be found in Supplementary file S3.

Sequencing and read processing

Sequencing was carried out by Eurofins Genomics Europe Sequencing GmbH (Jakob-Stadler-Platz 7, 78,467, Constance, GERMANY) on an Illumina MiSeq platform producing paired-read samples of 300 bp read length based on the V3-V4 amplicons. Raw sequences were quality controlled using CUTADAPT v2.7 [26] and used as input to QIIME2 v.2023.5 [27] on which they were denoised and clustered into ASVs (Amplicon Sequence Variants) using DADA2 [28]. ASV counts, sample metadata and taxonomy assignment tables are provided in Supplementary file S3.

Statistical analyses

All of the analyses were undertaken in R [29, 30]. Several different mathematical models with different forms for describing ISARs have been proposed but recent papers have underlined that the power model provides the best fit to the ISAR in most cases [11, 12]. Traditionally, the power model is linearized by logarithmic transformations of species richness S and area A using the equation $\log S = \log C + z \log A$, with z the slope of the resulting log–log relationship, and $\log C$ the intercept [11, 12]. Here, we employed the log–log linear modeling approach (using \log_{10}) to investigate the relationship between the number of Amplicon Sequence Variants (ASVs) detected per sample and tumor size. Prior to analysis, the completeness of each of the 27 samples was assessed using a rarefaction curve. As all rarefaction curves reach an asymptote, completeness was considered maximum for all samples and, consequently, with the number of ASVs directly comparable between tumors. However, as a sensitivity test, we refitted the ISAR log-log model using rarefied ASV richness using as sample size, the smallest library size encountered in our data i.e. 4231 sequences. Rarefaction was implemented using the R package *metag-Misc* [31] with the function *phyloseq_mult_raref_div* and the number of iterations set to 1000. For each model, three regression diagnostics were implemented, namely the Shapiro-Wilk test for the normality of the residuals, the Pearson's correlation between the fitted values and squared residuals for the homogeneity of the residuals, and the outlier t-test based on the Studentized residuals implemented with the function *outlierTest* in the *car* [32] R package. To assess the generality of our results, we adopted a repeated k -fold cross-validation approach

whereby we randomly partitioned the datasets into three equal components ($k=3$). For each partitioning, we put aside one component as the test data (9 samples) and fitted the log–log model to the remaining two components (the training data, 18 samples), and used the resultant model to predict the values of \log_{10} –species richness in the test data. The process was repeated for the three distinct combinations of training and test data. The predictive power of the log–log model was then assessed based on the Pearson's correlation calculated between the predicted and observed values of the test data and subsequently averaged across the three combinations. This 3-fold cross-validation process was then repeated 1000 times. This procedure was also implemented for the model fitted with rarefied species richness.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13099-024-00631-x>.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Author contributions

Conceptualization: K.A.T, N.D and M.G. Data Curation: N.D and M.G. Formal analysis: F.R and N.D; Funding acquisition: M.G; Investigation: All the authors contributed equally; Methodology: K.A.T, F.R, N.D; Supervision: K.A.T, M.G; Resources M.G, G.T, K.G; Writing – original draft: N.D, K.A.T, M.G, F.R, R.J.W; Writing – review & editing: All the authors contributed equally. All authors have approved the submission of this work.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethical approval

The research was approved by the Ethics committee of School of Medicine, National and Kapodistrian University of Athens, Greece (1586/27/1/14). The study was conducted in accordance with the declaration of Helsinki. All study participants signed an informed consent form.

Competing interests

The authors declare no competing interests.

Received: 21 March 2024 / Accepted: 28 June 2024

Published online: 07 July 2024

References

- West GB, Brown JH, Enquist BJ. A general model for the origin of allometric scaling laws in biology. *Science*. 1997;276(5309):122–6.
- Marquet PA, Quiñones RA, Abades S, Labra F, Tognelli M, Arim M, et al. Scaling and power-laws in ecological systems. *J Exp Biol*. 2005;208(9):1749–69.
- Masseran N. Power-law behaviors of the severity levels of unhealthy air pollution events. *Nat Hazards*. 2022;112(2):1749–66.
- Perkins TJ, Foxall E, Glass L, Edwards R. A scaling law for random walks on networks. *Nat Commun*. 2014;5(1):5121.
- Ritchie ME, Olff H. Spatial scaling laws yield a synthetic theory of biodiversity. *Nature*. 1999;400(6744):557–60.
- Lynch M, Conery JS. The origins of genome complexity. *Science*. 2003;302(5649):1401–4.
- Southwood TRE, May RM, Sugihara G. Observations on related ecological exponents. *Proceedings of the National Academy of Sciences*. 2006;103(18):6931–3.
- Locey KJ, Lennon JT. Scaling laws predict global microbial diversity. *Proc Natl Acad Sci*. 2016;113(21):5970–5.
- Dickey JR, Swenie RA, Turner SC, Winfrey CC, Yaffar D, Padukone A, et al. The utility of macroecological rules for microbial biogeography. *Front Ecol Evol*. 2021;9:633155.
- Ma Z. Sketching the human microbiome biogeography with DAR (diversity–Area relationship) profiles. *Microb Ecol*. 2019;77:821–38.
- Matthews TJ, Triantis KA, Whittaker RJ. The species–area relationship: theory and application. Cambridge University Press; 2021.
- Triantis KA, Guilhaumon F, Whittaker RJ. The island species–area relationship: biology and statistics. *J Biogeogr*. 2012;39(2):215–31.
- Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JL. The human microbiome project. *Nature*. 2007;449(7164):804–10.
- Sarmiento KR, Carr A, Diener C, Locey KJ, Gibbons SM. Island biogeography theory and the gut: why taller people tend to harbor more diverse gut microbiomes. *bioRxiv*. 2023.
- Brú A, Albertos S, Subiza JL, García-Asenjo JL, Brú I. The universal dynamics of tumor growth. *Biophys J*. 2003;85(5):2948–61.
- Banerjee J, Mishra N, Dhas Y, Metagenomics. A new horizon in cancer research. *Meta gene*. 2015;5:84–9.
- Luo W. Nasopharyngeal carcinoma ecology theory: Cancer as multidimensional spatiotemporal unity of ecology and evolution pathological ecosystem. *Theranostics*. 2023;13(5):1607.
- Chroni A, Kumar S. Tumors are evolutionary island-like ecosystems. *Genome Biol Evol*. 2021;13(12):evab276.
- Chen Y, Wu F-H, Wu P-Q, Xing H-Y, Ma T. The role of the tumor microbiome in tumor development and its treatment. *Front Immunol*. 2022;13:935846.
- Yang L, Li A, Wang Y, Zhang Y. Intratumoral microbiota: roles in cancer initiation, development and therapeutic efficacy. *Signal Transduct Target Therapy*. 2023;8(1):35.
- Rezasoltani S, Aghdaei HA, Jasemi S, Gazouli M, Dovrolis N, Sadeghi A, et al. Oral microbiota as novel biomarkers for Colorectal Cancer Screening. *Cancers*. 2022;15(1):192.
- Galeano Niño JL, Wu H, LaCourse KD, Kempchinsky AG, Baryjames A, Barber B, et al. Effect of the intratumoral microbiota on spatial and cellular heterogeneity in cancer. *Nature*. 2022;611(7937):810–7.
- Murgai M, Giles A, Kaplan R. Physiological, tumor, and metastatic niches: opportunities and challenges for targeting the tumor microenvironment. *Crit Reviews™ Oncog*. 2015;20:3–4.
- Gatenbee CD, Minor ES, Slebos RJ, Chung CH, Anderson AR. Histoecology: applying ecological principles and approaches to describe and predict tumor ecosystem dynamics across space and time. *Cancer Control*. 2020;27(3):1073274820946804.
- Chen J, Li T, Liang J, Huang Q, Huang J-D, Ke Y, et al. Current status of intratumour microbiome in cancer and engineered exogenous microbiota as a promising therapeutic strategy. *Biomed Pharmacother*. 2022;145:112443.
- Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*. 2011;17(1):10–2.
- Estaki M, Jiang L, Bokulich NA, McDonald D, González A, Kosciolk T, et al. QIIME 2 enables comprehensive end-to-end analysis of diverse microbiome data and comparative studies with publicly available data. *Curr Protocols Bioinf*. 2020;70(1):e100.
- Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016;13(7):581–3.
- R Core Team R. R: A language and environment for statistical computing. 2023.
- Ihaka R, Gentleman R. R: a language for data analysis and graphics. *J Comput Graphical Stat*. 1996;5(3):299–314.
- Mikryukov V. vmikr/metagMisc: v.0.0.0.9000 (v.0.0.0.9000): Zenodo; 2017. <https://doi.org/10.5281/zenodo.571403>.
- Fox J, Weisberg S. An R companion to applied regression. Sage; 2011.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.