**Codon bias in ribosomal protein genes associated with transcription and growth in soil microbial communities during rewetting.**

Peter F. Chuckran, Nikita Goudner (maybe), Jeff Kimbrell, Katerina Estera-Molina, Alexa M. Nicolas, Ella T. Sieradzki, Jillian Banfield (if she wants), Paul Dijkstra, Mary K. Firestone, Jennifer Pett-Ridge, Steve J. Blazewicz

ABSTRACT

Genomic traits are informative metrics for assessing life-strategies of soil microorganisms; however, linking these traits to growth has remained elusive. In soil bacteria, traits such as high codon usage and small genome size have been linked to fast growth, yet often through indirect measurements or modeling approaches. Here, we use stable isotope probing with 18O-water to track growth of soil microorganisms over the course of one week following rewetting after a prolonged dry period—a phenomenon that is both a massive disturbance and a period of high microbial activity. By mapping the isotopically enriched DNA fractions against metagenome-assembled-genomes (MAGs), we linked taxa-specific growth rates with genomic traits. We found that fast growing bacterial taxa did not have a greater relative abundance of ribosomal proteins or transcriptional regulators, but rather high levels of codon bias in ribosomal protein genes, and relationships between GC content and growth appeared to be largely dependent on codon bias. Through mapping metatranscriptomes against these MAGs, we found higher upregulation of ribosomal protein genes that had high levels of codon usage bias. Although faster growing organisms tended to have smaller genomes on average, this trait only emerged after a week of growth—suggesting codon usage bias as the predominant trait for rapid transcription and growth after disturbance. [Add solid closing sentence here]

INTRODUCTION

In ecosystems characterized by dry and wet seasons, the first rain event following the dry season results in the rapid mineralization of soil carbon. This phenomenon, wherein the rewetting of dry soil releases a pulse of CO2, is known as the “Birch Effect” (1), and can account for a large portion of the annual soil respiration in systems driven by dry and wet seasons (2). The rapid stimulation of activity is driven by both the release of water stress as well as an influx in the bioavailability of C compounds sourced from osmolytes (3–5), microbial necromass (6, 7), slaking of microaggregates (8), and increased connectivity in the soil matrix (9, 10). The microbial response to rewetting therefore has important consequences for the stabilization of soil organic matter; however, we lack an understanding of which traits control the growth and activity of this response.

Several traits are known to influence or relate to the growth rate and activity of microorganisms. These can be broken up into two broad categories. First is the abundance of specifical functional genes, such as those encoding for ribosomal proteins, rRNA genes, and transcription factors, which have been shown to be associated with higher maximum growth rate. *Maybe a little bit about the evidence for this.*

Second are the traits that describe genes and genomes, such as nucleotide frequency, codon usage, and genome (or gene) size. Codon usage in particular has been shown to be a strong determinant of both transcription and growth rate. A high alignment of gene codon frequencies to the anticodons of the tRNA pool increases the rate of both transcription and translation (11)—commonly referred to as codon optimization. The level of codon optimization of a transcript has important consequences for translation; affecting the rate of elongation, protein folding, initiation, and termination (12–14). Codon optimization is also a strong determinant of mRNA abundance. A high abundance of optimized codons generally increases mRNA stability (15, 16) and high codon optimization has also been shown to be independently related to higher levels of transcription (17, 18). “Codon bias” describes codon redundancy in the genetic code for a particular gene or genome, and generally correlates to the degree of codon optimization. High levels of codon bias in ribosomal proteins is associated with rapid growth in bacteria (19, 20) and has been increasingly used as a predictor of maximum growth rate (18–20)(21, 22).

The selection of specific nucleotides and amino acids can additionally alter gene expression through differential requirement for energy and nutrients. For example, thymine (T) requires more energy to synthesize then adenine (A), cytosine (C) requires more energy than guanine (G), and G+C are more costly than A+T (23). For this reason, highly upregulated genes tend to have a lower AT and GC skew (i.e. more A than T, more C than G; Chen et al., 2016). However, despite the higher cost of synthesis, a high GC content has been associated with high levels of gene expression in bacteria. There are likely several reasons for this, one of which being that high GC codons tend to encode for biosynthetically cheaper amino acids which often have a higher level of expression (24).

Genome size is another commonly cited genomic trait theorized to be associated with growth rate in free living bacteria. Most often, this is presented in the context of a copiotroph-oligotroph framework, where bacteria in low nutrient environments tend to have smaller genomes and slower growth rates (25)—such as in the case of genomic streamlining (26). However, it has also been hypothesized that soils might have larger genomes for greater metabolic diversity at the expense of growth rate (27). Although it has been shown that genome size may relate to life strategy and soil environment (28–30), several studies have found no relationship between genome size and growth rate in free living bacteria (31–33).

In soils, a major hurdle in assessing how traits relate to growth rate has been our inability to effectively measure these traits ­*in situ*. Much of the work linking genomic traits to growth in soils has been conducted using isolate data from databases where the growth rate is often assessed from culture. It has been shown that traits associated with growth of microbes grown on media might not translate to free-living microbes in their natural environments (34), making it difficult to rely on these trends when assessing the functional potential of soil bacteria. Fortunately, recent methodological and technological advancements have created new opportunities to track the growth of microbes *in situ*. Quantitative stable isotope probing (qSIP) uses the incorporation of added stables isotopes, such as 18O-water, in DNA to track the growth rate of microbes based on sequence (35). Although this approach has mostly been conducted with amplicon sequencing, recent work conducted with metagenomes has provided valuable insights into the genes associated with growth (7, 36, 37).

In this study, we pair transcription data and growth rates of metagenomes assembled genomes in order to assess how traits are associated with activity and growth during the rewetting of dry soil at the end of the dry season in a Mediterranean grassland. Rewetting represents an ecologically important phenomenon where traits associated with growth and activity may play an important role for stabilizing soil carbon. Here, we investigate how traits such as genome size, nucleotide selection, codon usage, and ribosomal and transcription factor gene frequency, relate to growth and transcription. Linking these traits to activity may not only provide insight into the mechanisms of the Birch Effect, but also give valuable insight into the forces underpinning the selection and evolution of soil microorganisms.

METHODS

The initial field study from which soils were collected was located at Hopland Extension and Research Center (HERC), Hopland, California, USA (39.00092, -123.07962), which resides on the ancestral home of the Shóqowa and Hopland people. The region has a Mediterranean climate, with warm dry summers and cool wet winters. The field site was located in a field dominated by *Avena barbata* (wild oat grass). This study consisted of 16 plots 3.24 m2 in a randomized block design across X. Rain out shelters were constructed around each plot, either allowing full or 50% precipitation. A full description of the field experimental set-up can be found in *[Rina’s future paper]*

Soils were collected in August 2018 before the first rain event at the end of the dry season. At the time of collection the soil gravimetric water content was approximately 3%. Soil was taken from 8 randomly selected plots (4 full and 4 reduced precipitation) with 0-15 cm cores. Samples were transported to Lawrence Livermore National Laboratory where they were pooled by plot, run through a 2 mm sieve, and picked of roots and large debris.

*Wet-up experiment*

Details of the H218O labeling can be found in Nicolas et al., 2022 and Sieradzki et al., 2022. Soils from each plot were separated into 12 chambers containing 5 g of soil each. Soils were brought to 22% gravimetric water content—half with natural abundance water and the other half with 98-at% H218O. A pair of labeled and unlabeled samples from each plot were destructively harvested at 0, 3, 24, 48, 72, and 168 h post wet-up. Samples were immediately frozen in liquid-N2 and stored in a -80°C freezer.

*Metagenomic qSIP*

A full description of the DNA extraction, sequencing, assembly, binning, and qSIP calculations can be found in Sieradzki et al., 2022. Briefly, DNA was extracted in triplicate using a phenol chloroform extraction and then pooled. Samples were then spun on an ultracentrifuge in a cesium chloride solution to create a density gradient. The contents of the ultracentrifuge tube was then separated into 36 fractions, each of which were then assessed for density and the DNA purified and concentrated. Fractions were combined into 5 groups along the density gradient and sequenced on an Illumina Novaseq platform.

Adapters were trimmed and reads were QC filtered using bbduk (38) and Sickle (39). QC filtered reads were assembled into contigs using MEGAHIT (v1.2.9; (40), binned with MaxBin 2.0 (41) and MetaBAT2 (42), and refined with MetaWrap (43). We combined this set of MAGs with another set binned from a study at a nearby site at HERC (36) and dereplicated MAGs using dRep (44). Reads from each fraction were then mapped against the set of MAGs using BBMap (38). DRAM (45) was used to identify open reading frames (ORFs) using Prodigal (46) and annotate functional genes against the KEGG database (47).

*RNA Extraction and sequencing*

Since no isotopic work was conducted with the RNA samples, we randomly chose 4 samples from each precipitation treatment to track RNA transcription at all timepoints. RNA was extracted from soils using a RNeasy PowerSoil Total RNA kit (Qiagen) according to manufacturer instructions. Extracted RNA was treated with RNase-free DNase (Qiagen) and stored at -80°C. RNA concentration was determined using a Qubit fluorometer (Invitrogen, Carlsbad, CA, USA) and quality was assessed using a Nanodrop One Spectrophotometer (Thermo Scientific Invitrogen). Samples were sent to the Joint Genome Institute (JGI; Berkeley, California, USA), for sequencing. Paired-end 2 x 151 bp libraries were sequenced on an Illumina NovaSeq platform. Metatranscriptomes were run through the JGI Integrated Microbial Genomes (IMG) pipeline v.5.1.5 (48) and can be found under the GOLD project ID Gp0612223. Although IMG assemblies are not included in this analysis, a more detailed description can be found in the data release (DOI: and citation).

*Metatranscriptome analysis*

Raw reads were downloaded from the JGI genome portal in September 2022. Reads were QC filtered using bbduk (38) and Sickle (39). BBmap was then used to map QC filtered reads to a concatenated reference mag file (minid=0.95). FeatureCounts (49) was then used to extract the number of reads mapped to genes identified by DRAM. Expression and normalization values of mapped transcripts for each annotated gene (excluding rRNA genes) were generated using DESeq2 (50). Nonannotated genes were not included in the normalization, as some of these may represent rRNA genes—the abundance of which would be representative of the efficiency of rRNA depletion, as opposed to any biologically meaningful mechanism.

*Genomic traits*

Genome size was calculated for medium to high quality mags by multiplying the total base pairs assembled by the completeness of the genome. For each gene in each MAG we calculated the nucleotide frequencies and effective number of codons considering the genomic background nucleotide frequency (ENC’), as described by *Novembre 2002* (52). We also calculated the ENC’ of the ribosomal proteins, which was used to calculate ΔENC’ (20):

Where ENC’all is the ENC’ for the whole genome, and ENC’ribo is the ENC’ of the ribosomal protein genes. Higher ΔENC’ values represent greater codon bias in ribosomal protein genes.

RESULTS

Since single copy genes may be a misleading predictor of genome completeness (51), we assessed genome completeness as it relates to growth rate and found little relationship. Although this

DISCUSSION

To include:

* Even though our results show that small genomes have a higher growth rate, this is not necessarily to say that smaller genomes have a higher maximum growth rate overall, but rather that growth under these conditions favors smaller genomes. For example, subtle wetting with 18O water in a tropical soil during the wet season might show the opposite result. However, the results here indicate that the response to rewetting potentially favors smaller genomes that can respond rapidly to change, perhaps at the consequence of metabolic diversity. This might be why we find smaller genomes in arid areas. Studies have shown that smaller genomes tend to be associated with high pH and increased aridity, and the ability to respond to short term pulses of labile carbon might contribute to this relationship.

REFERENCES

1. H. F. Birch, The effect of soil drying on humus decomposition and nitrogen availability. *Plant and Soil* **10**, 9–31 (1958).

2. P. Jarvis, *et al.*, Drying and wetting of Mediterranean soils stimulates decomposition and carbon dioxide emission: the “Birch effect.” *Tree Physiology* **27**, 929–940 (2007).

3. N. Fierer, J. P. Schimel, A Proposed Mechanism for the Pulse in Carbon Dioxide Production Commonly Observed Following the Rapid Rewetting of a Dry Soil. *Soil Science Society of America Journal* **67**, 798 (2003).

4. E. W. Slessarev, *et al.*, Cellular and extracellular C contributions to respiration after wetting dry soil. *Biogeochemistry* **147**, 307–324 (2020).

5. C. R. Warren, Response of osmolytes in soil to drying and rewetting. *Soil Biology and Biochemistry* **70**, 22–32 (2014).

6. S. J. Blazewicz, E. Schwartz, M. K. Firestone, Growth and death of bacteria and fungi underlie rainfall-induced carbon dioxide pulses from seasonally dried soil. *Ecology* **95**, 1162–1172 (2014).

7. A. M. Nicolas, *et al.*, Isotope-enrichment reveals active viruses follow microbial host dynamics during rewetting of a California grassland soil. 2022.09.30.510406 (2022).

8. P. M. Homyak, *et al.*, Effects of altered dry season length and plant inputs on soluble soil carbon. *Ecology* **99**, 2348–2362 (2018).

9. F. E. Moyano, S. Manzoni, C. Chenu, Responses of soil heterotrophic respiration to moisture availability: An exploration of processes and models. *Soil Biology and Biochemistry* **59**, 72–85 (2013).

10. A. P. Smith, *et al.*, Shifts in pore connectivity from precipitation versus groundwater rewetting increases soil carbon loss after drought. *Nature Communications* **8** (2017).

11. J. B. Plotkin, G. Kudla, Synonymous but not the same: The causes and consequences of codon bias. *Nature Reviews Genetics* **12**, 32–42 (2011).

12. Y. Liu, Q. Yang, F. Zhao, Synonymous but Not Silent: The Codon Usage Code for Gene Expression and Protein Folding. *Annual Review of Biochemistry* **90**, 375–401 (2021).

13. C. H. Yu, *et al.*, Codon Usage Influences the Local Rate of Translation Elongation to Regulate Co-translational Protein Folding. *Molecular Cell* **59**, 744–754 (2015).

14. M. Zhou, *et al.*, Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature 2013 495:7439* **495**, 111–115 (2013).

15. C. Dressaire, *et al.*, Role of mRNA Stability during Bacterial Adaptation. *PLOS ONE* **8**, e59059 (2013).

16. V. Presnyak, *et al.*, Codon Optimality Is a Major Determinant of mRNA Stability. *Cell* **160**, 1111–1124 (2015).

17. Z. R. Newman, J. M. Young, N. T. Ingolia, G. M. Barton, Differences in codon bias and GC content contribute to the balanced expression of TLR7 and TLR9. *Proceedings of the National Academy of Sciences of the United States of America* **113**, E1362–E1371 (2016).

18. Z. Zhou, *et al.*, Codon usage is an important determinant of gene expression levels largely through its effects on transcription. *Proceedings of the National Academy of Sciences of the United States of America* **113**, E6117–E6125 (2016).

19. D. L. Kirchman, Growth rates of microbes in the oceans. *Annual Review of Marine Science* **8**, 285–309 (2016).

20. S. Vieira-Silva, E. P. C. Rocha, The Systemic Imprint of Growth and Its Uses in Ecological (Meta)Genomics. *PLoS Genetics* **6**, e1000808 (2010).

21. A. M. Long, S. Hou, • J Cesar, I. -Espinoza, J. A. Fuhrman, Benchmarking microbial growth rate predictions from metagenomes. *The ISME Journal*, 1–13 (2020).

22. J. L. Weissman, S. Hou, J. A. Fuhrman, Estimating maximal microbial growth rates from cultures, metagenomes, and single cells via codon usage patterns. *Proceedings of the National Academy of Sciences* **118**, e2016810118 (2021).

23. W.-H. Chen, G. Lu, P. Bork, S. Hu, M. J. Lercher, Energy efficiency trade-offs drive nucleotide usage in transcribed regions. *Nature Communications* **7**, 11334 (2016).

24. D. W. Raiford, *et al.*, Metabolic and translational efficiency in microbial organisms. *Journal of Molecular Evolution* **74**, 206–216 (2012).

25. F. M. Lauro, *et al.*, The genomic basis of trophic strategy in marine bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 15527–33 (2009).

26. S. J. Giovannoni, J. Cameron Thrash, B. Temperton, Implications of streamlining theory for microbial ecology. *The ISME Journal* **8**, 1553–1565 (2014).

27. K. T. Konstantinidis, J. M. Tiedje, Trends between gene content and genome size in prokaryotic species with larger genomes. **10**, 3160–3165 (2004).

28. S. Beier, J. Werner, T. Bouvier, N. Mouquet, C. Violle, Trait-trait relationships and tradeoffs vary with genome size in prokaryotes. *Frontiers in Microbiology* **13** (2022).

29. Y. Chen, J. W. Neilson, P. Kushwaha, R. M. Maier, A. Barberán, Life-history strategies of soil microbial communities in an arid ecosystem. *ISME Journal* **15**, 649–657 (2021).

30. P. F. Chuckran, *et al.*, Edaphic controls on genome size and GC content of bacteria in soil microbial communities. *Soil Biology and Biochemistry* **178**, 108935 (2023).

31. S. Freilich, *et al.*, Metabolic-network-driven analysis of bacterial ecological strategies. *Genome Biology* **10**, R61 (2009).

32. S. Vieira-Silva, M. Touchon, E. P. C. Rocha, No evidence for elemental-based streamlining of prokaryotic genomes. *Trends in Ecology and Evolution* **25**, 319–320 (2010).

33. M. Westoby, *et al.*, Cell size, genome size, and maximum growth rate are near-independent dimensions of ecological variation across bacteria and archaea. *Ecology and Evolution* **11**, 3956–3976 (2021).

34. J. Li, *et al.*, Predictive genomic traits for bacterial growth in culture versus actual growth in soil. *The ISME Journal*, 1 (2019).

35. B. A. Hungate, *et al.*, Quantitative microbial ecology through stable isotope probing. *Applied and Environmental Microbiology* **81**, 7570–7581 (2015).

36. A. Greenlon, *et al.*, Quantitative Stable-Isotope Probing (qSIP) with Metagenomics Links Microbial Physiology and Activity to Soil Moisture in Mediterranean-Climate Grassland Ecosystems. *mSystems* **7**, e00417-22 (2022).

37. E. T. Sieradzki, *et al.*, Functional succession of actively growing soil microorganisms during rewetting is shaped by precipitation history. *bioRxiv*, 2022.06.28.498032 (2022).

38. B. Bushnell, BBTools Software Package (2014).

39. Joshi NA, Fass JN., Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) (2023) (May 11, 2023).

40. D. Li, C. M. Liu, R. Luo, K. Sadakane, T. W. Lam, MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).

41. Y.-W. Wu, B. A. Simmons, S. W. Singer, MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605–607 (2016).

42. D. D. Kang, *et al.*, MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).

43. G. V. Uritskiy, J. DiRuggiero, J. Taylor, MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **6**, 158 (2018).

44. M. R. Olm, C. T. Brown, B. Brooks, J. F. Banfield, dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J* **11**, 2864–2868 (2017).

45. M. Shaffer, *et al.*, DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Research* **48**, 8883–8900 (2020).

46. D. Hyatt, *et al.*, Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 1–11 (2010).

47. M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, K. Morishima, KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research* **45**, D353–D361 (2017).

48. I.-M. A. Chen, *et al.*, IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Research* **47**, D666–D677 (2019).

49. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

50. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15** (2014).

51. L.-X. Chen, K. Anantharaman, A. Shaiber, A. Murat Eren, J. F. Banfield, Accurate and complete genomes from metagenomes (2020) https:/doi.org/10.1101/gr.258640.119 (July 29, 2020).

52. J. A. Novembre, Accounting for Background Nucleotide Composition When Measuring Codon Usage Bias. *Molecular Biology and Evolution* **19**, 1390–1394 (2002).