Bacterial natural product biosynthetic domain composition in soil correlates with changes in latitude on a continent-wide scale

Christophe Lemetre[a,1], Jeffrey Maniko[a,1], Zachary Charlop-Powers[a], Ben Sparrow[b,c], Andrew J. Lowe[c], and Sean F. Brady[a,2]

[a]Laboratory of Genetically Encoded Small Molecules, The Rockefeller University, New York, NY 10065; [b]Terrestrial Ecosystems Research Network, University of Adelaide, Adelaide, SA 5005, Australia; and [c]Environment Institute and School of Biological Sciences, University of Adelaide, Adelaide, SA 5005, Australia

Although bacterial bioactive metabolites have been one of the most prolific sources of lead structures for the development of small-molecule therapeutics, very little is known about the environmental factors associated with changes in secondary metabolism across natural environments. Large-scale sequencing of environmental microbiomes has the potential to shed light on the richness of bacterial biosynthetic diversity hidden in the environment, how it varies from one environment to the next, and what environmental factors correlate with changes in biosynthetic diversity. In this study, the sequencing of PCR amplicons generated using primers targeting either ketosynthase domains from polyketide biosynthesis or adenylation domains from nonribosomal peptide biosynthesis was used to assess biosynthetic domain composition and richness in soils collected across the Australian continent. Using environmental variables collected at each soil site, we looked for environmental factors that correlated with either high overall domain richness or changes in the domain composition. Among the environmental variables we measured, changes in biosynthetic domain composition correlate most closely with changes in latitude and to a lesser extent changes in pH. Although it is unclear at this time the exact mix of factors that may drive the relationship between biosynthetic domain composition and latitude, from a practical perspective the identification of a latitudinal basis for differences in soil metagenome biosynthetic domain compositions should help guide future natural product discovery efforts.

chemical biogeography | eDNA | polyketide synthase | nonribosomal peptide synthetase | continental soil analysis

Bacteria rely heavily on small molecules (natural products, NPs) to interact with their environment (1). Among other roles, NPs serve as chelators, toxins, and signaling agents that allow bacteria to obtain nutrients as well as defend against, and communicate with, other organisms. In addition to their natural ecological roles, bacterial secondary metabolites are of interest as they have served as one of the most productive sources of lead structures for developing small-molecule therapeutics. Understanding how NP biosynthetic gene clusters are distributed in the environment and what factors drive their distribution is therefore of interest both from an ecological perspective and from a practical therapeutic discovery perspective. It is now well established that most bacteria present in the environment are not readily cultured in the laboratory (2–4); as a result, the NPs they produce remain hidden in the environment, thus preventing the comparative analysis of secondary metabolomes from different environments using culture-based analytical chemistry methods. The sequencing of DNA extracted directly from environmental samples (environmental DNA, eDNA) can be used to estimate bacterial biosynthetic domain richness in an environment and to gauge how biosynthetic domain composition varies from one environment to the next. In this context, richness is intended to be a measure of the sheer number of different sequences predicted to be present in an environment, while composition reflects the type of sequences present. Unfortunately, it remains impractical to shotgun-sequence complex environmental

microbiomes to a depth that can reliably yield information about NP biosynthetic diversity. In a process similar to the use of 16S genes amplified from environments to profile bacterial species diversity, bacterial biosynthetic diversity can be assessed using degenerate primers targeting conserved genes found across large NP biosynthetic families. Although bacterial NPs comprise an amazing diversity of chemical structures, many of the most biomedically interesting metabolites arise from only a small number of evolutionarily related biosynthetic families. Two of the most common classes of NPs that have yielded biomedically interesting metabolites arise from polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) biosynthetic gene clusters (5). In a growing number of studies, individual next-generation sequencing reads (NP sequence tags, NPSTs) derived from PCR amplicons generated using primers targeting domains in PKS or NRPS gene clusters have been used to predict the biosynthetic richness and diversity present in environmental metagenomes (6–12). Unfortunately, the ad hoc nature of existing domain sequencing studies has made it difficult to identify environmental features associated with differences in either biosynthetic domain composition or richness.

To more systematically identify factors that correlate with changes in biosynthetic composition we sought to assess NRPS adenylation (AD) domain and PKS ketosynthase (KS) domain diversity across a well-characterized collection of soils derived from a large and ecologically diverse land mass. The Australian

---

Significance

Secondary metabolites encoded by soil bacteria have proved to be a rich source of lead structures for the development of small-molecule therapeutics. Environmental factors that contribute to differences in the composition of natural product biosynthetic gene clusters found in geographically distant soil microbiomes have remained elusive. In this study, we sought to address this key outstanding question. Of the factors we assessed, changes in latitude were found to correlate most consistently with changes in biosynthetic domain composition on a continent-wide scale. Although further studies are needed to better understand the underlying causes driving this relationship, these findings provide insights into how best to direct future natural product drug discovery efforts.

---

[1]C.L. and J.M. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: sbrady@rockefeller.edu.

CHEMISTRY

MICROBIOLOGY

continent provides a unique natural laboratory for this type of large-scale study, as it is both an isolated and ecologically diverse landmass (13). Here we report on KS and AD domain composition and richness in Australian soil on a continent-wide scale. In this analysis, the most consistent correlations were observed between changes in biosynthetic domain composition and changes in latitude and, to a lesser extent, pH, suggesting that both edaphic and nonedaphic factors may be important in determining the biosynthetic gene cluster content of environmental microbiomes. A growing body of evidence from whole-genome sequencing studies suggests that bacterial speciation is, at least in part, likely to depend on the differentiation of an organism's secondary metabolome (14–17). As a consequence, understanding secondary metabolism distribution patterns has the potential not only to assist with the practical search of novel bioactive NPs but also to inform on factors that determine species distribution in nature,

one of the key outstanding questions in the fields of ecology and biogeography (18).

Results and Discussion

Sample Collection and Sequencing. As part of an ongoing continent-wide environmental monitoring program (AusPlots), the Terrestrial Ecosystem Research Network (TERN) group at the University of Adelaide collected topsoil from 397 ecologically and geographically diverse sites across Australia; 361 samples were collected from areas broadly defined as rangeland, with the remaining 36 samples collected from northern and southern coastal environments. Rangeland, colloquially known as the "outback," makes up ~81% of Australia. The rangelands consist of a diverse collection of largely unmodified ecosystems that include shrublands, grasslands, woodlands, and tropical savannas, which together make up many of the habitats seen across Australia
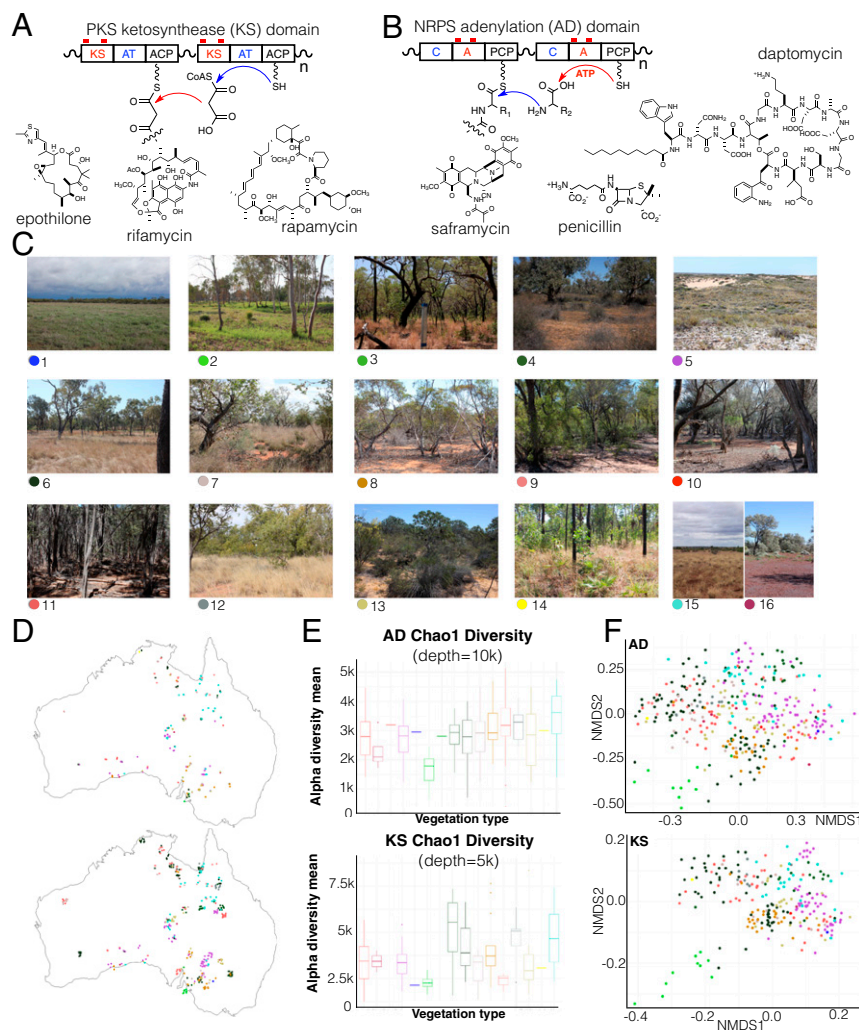

Fig. 1. Biosynthesis and ecotype: (A) Polyketide and (B) nonribosomal peptide assembly line-like biosynthesis is responsible for producing a disproportionately large fraction of biomedically relevant bacterial NPs. For this analysis we used degenerate primers targeting AD and KS domains to assess biosynthetic diversity in Australian soils. (C) Representative pictures of the 16 different ecotypes from which soil samples were collected (1: cleared, nonnative vegetation, buildings; 2: eucalyptus low open forests; 3: eucalyptus open forests; 4: eucalyptus open woodlands; 5: Chenopod shrublands, Samphire shrublands, and orblands; 6: eucalyptus woodlands; 7: Mallee open woodlands and sparse Mallee shrublands; 8: Mallee woodlands/shrublands; 9: Melaleuca forests and woodlands; 10: Casuarina forests and woodlands; 11: Acacia forests and woodlands; 12: other forests and woodlands; 13: other shrublands; 14: tropical eucalyptus woodlands/grasslands; 15: tussock grasslands; 16: Acacia shrublands). (D) Collection sites in Australia are color-coded according to ecotype. Because of the scale of the map, some collection sites that appear to be marked with a single marker area actually indicative of more than one sample being collected (upper map). On the lower map, individual sites on the map were jittered using R's jitter function. All subsequent sample site maps are presented with a jittering. (E) KS and AD domain Chao1 diversity estimates by sample site ecotype. (F) AD and KS domain NMDS ordination plots are color-coded according to sample-site ecotype.

(13). Our sampling efforts were a pragmatic balance between covering diverse ecotypes and geographies and being logistically limited by the remote regions of northwestern Australia. We aimed to collect multiple samples from the major vegetation groups shown in Fig. 1 (see also Figs. S1 and S2). The samples used in this study encompass much of the diversity of soil types found in Australia. Each collection site was assigned to a major vegetation group based upon the structure and composition of plants observed at that site. Latitude, longitude, altitude, and soil pH were recorded at each site, and average yearly rainfall and yearly average daily high temperature were obtained for each site from modeled data (19, 20).

To compare and contrast biosynthetic domain diversity, eDNA was extracted from soil collected at each site (96-well PowerSoil-HTP soil DNA isolation kit; MoBio). NRPS AD and PKS KS domains were then PCR-amplified from each eDNA isolate using domain-specific degenerate primers (21, 22). PCR amplification with degenerate primers using crude eDNA as a template is not always successful. In this study, we were able to obtain amplicons from 246 sites using degenerate KS primers and 367 sites using degenerate AD primers. Forward and reverse primers contained different collections of 8-bp barcodes resulting in amplicons from each site containing unique 16-bp barcodes upon concatenation of the two 8-bp barcodes. Barcoded amplicons generated from all

soils were pooled and sequenced using Illumina MiSeq technology (2 × 300 bp). Forward and reverse reads were concatenated together, and the KS and AD operational taxonomic units (OTUs) present in each site were generated using Usearch by clustering the reads at 95% sequence identity (23). OTU tables generated from the site-specific amplicon sequence data in conjunction with metadata recorded at each collection site were used to compare and contrast NP biosynthetic domain composition and richness across the Australian continent.

Factors That Correlate with Changes in Biosynthetic Domain Richness on a Continent-Wide Scale. KS and AD domain OTU richness at each collection site were estimated using the Chao1 diversity metric. When subsampled to an even depth of 5,000 and 10,000 reads for AD and KS data, respectively, AD domain estimates ranged from below 1,000 to above 5,000 OTUs per soil and KS domain estimates ranged from ∼2,000 to over 7,500 OTUs per soil. Despite this variation in sample-to-sample richness, we detected only minimal correlations between biosynthetic domain richness and any environmental factor we measured (Fig. 1). Interestingly, our data also did not show a latitudinal gradient of biosynthetic domain richness across the Australia continent (Figs. S3–S5), whereas such a gradient of increased species richness from the poles to the equator is well established in the ecology of
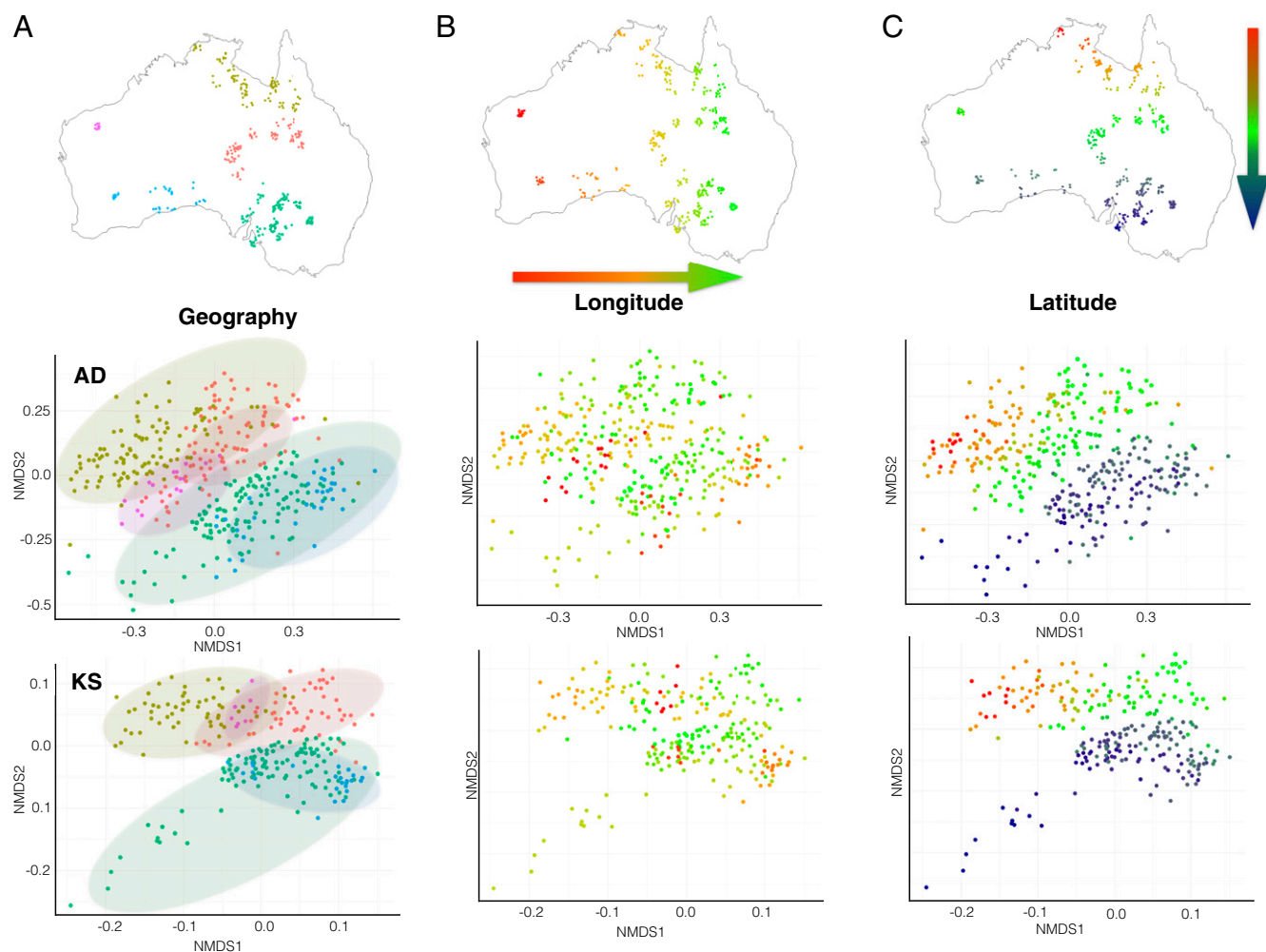

Fig. 2. Geographic parameters. (A) Australian collection sites were grouped into five groups by geographic proximity (North, central east, central west, southwest, and southeast). NDMS ordination plot for AD and KS domains color-coded based on the geographic proximity. Samples sites and NDMS ordination plot for AD and KS domains are color-coded according to either longitude (B) or latitude (C).

animals and plants (24–27). Whether a biosynthetic domain richness gradient exists on a larger global scale will require future analysis of biosynthetic domains across a wider set of latitudes.

Factors That Correlate with Changes in Biosynthetic Domain Composition on a Continent-Wide Scale. Relationships between the populations of biosynthetic domain sequences from different soils were explored using ordination analysis of the AD and KS OTU tables. After removing very rare OTUs (present at two or fewer sites) and sites containing fewer than 1,000 reads, we performed an ordination analysis of the OTU tables using nonmetric multidimensional scaling (NMDS) (28). To identify factors that most strongly correlate with site-to-site differences in biosynthetic domain composition,

environmental variables associated with each site were mapped onto the resulting KS and AD NMDS ordination plots (Figs. 2 and 3).

Both KS and AD NMDS ordination plots were initially annotated according to the 16 different vegetation descriptors used to classify collection sites (Fig. 1). Within the resulting ecology-annotated NMDS ordination plots we only observed vegetation type clustering when all samples with the same ecotype descriptor were obtained within close geographic proximity to each other (Fig. 1 and Figs. S6 and S7). In cases where the samples were collected from geographically distant sites but classified as the same ecotype we observed clustering based on geography, suggesting that differences in vegetation do not directly govern differences in bacterial biosynthetic composition.
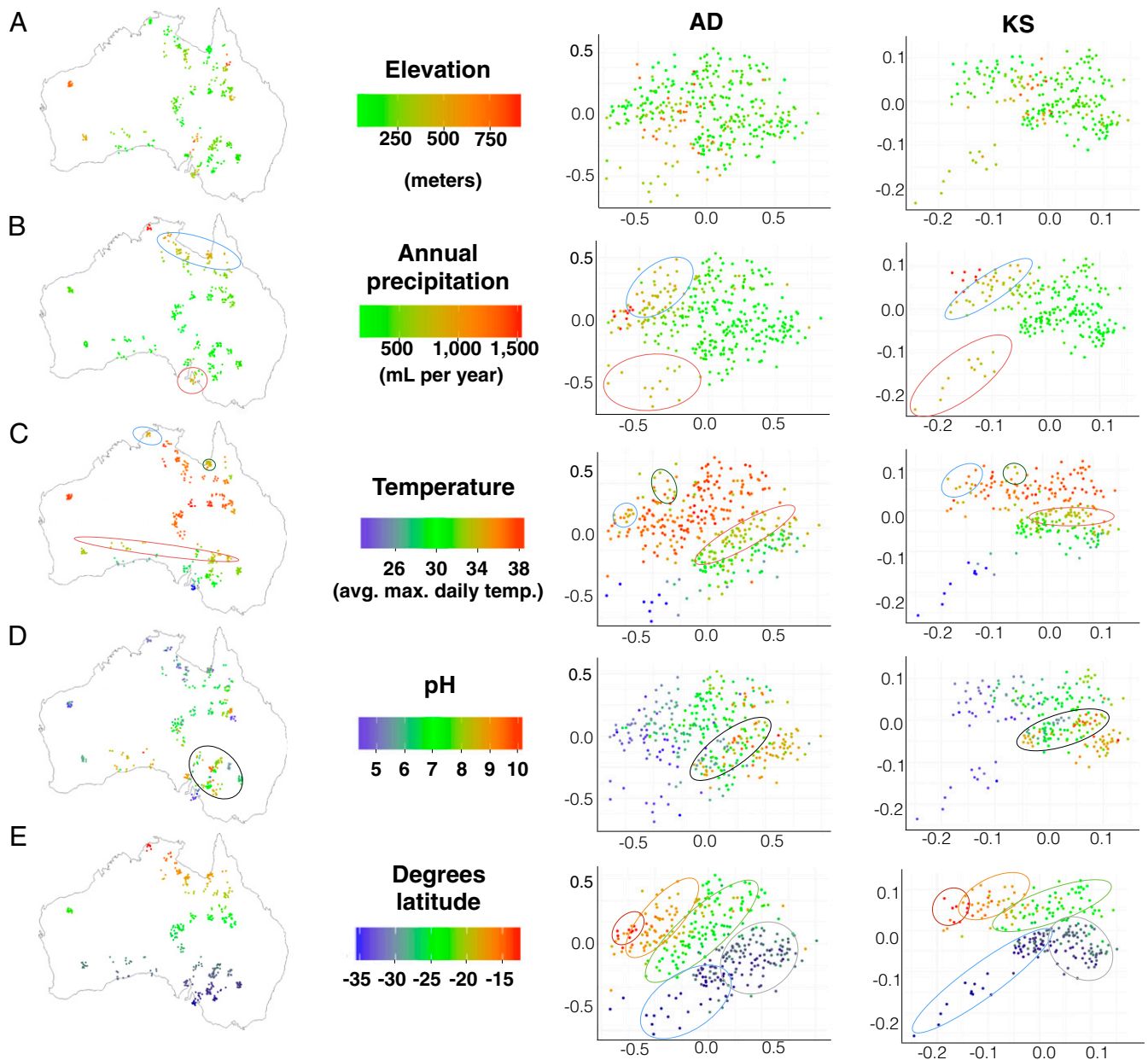

Fig. 3. NDMS ordination plot metadata correlations: Sample site map and KS and AD domain NMDS ordination plots color-coded according to (*A*) elevation, (*B*) annual precipitation, (*C*) temperature, (*D*) pH, or (*E*) latitude. In *B* and *C*, the colored ovals highlight samples that appear at distinct regions on the NMDS ordination plots but have similar environment variable values. In *D*, the black oval highlights the clusters of samples with a wide range of pH in a similar region of the NMDS plot. In *E*, the ovals highlight the clustering of samples collected at similar latitudes. If environmental data are missing at a site, the corresponding point is hidden in the NMDS plot.

To look for potential geographic relationships in the biosynthetic domain composition data, the 397 collection sites were initially divided into five general geographical regions [Fig. 2; north (yellow), central west (purple), central east (red), southwest (blue), and southeast (green)]. When KS and AD domain NMDS ordination plots were annotated according to these gross geographic groupings, general geographic clustering patterns emerged in both the KS and AD ordination plots (Fig. 2). To further explore a potential geographic basis for differences in biosynthetic domain composition, NMDS ordination plots were labeled according to either collection-site latitude or longitude. Regardless of a sample's ecological origin, we observed a continent-wide correlation between a soil's latitudinal position and its NMDS ordination plot-defined biosynthetic composition.

Sample-site elevation, precipitation, average temperature (i.e., nonedaphic factors), and pH (i.e., edaphic factor) were recorded for each soil collection site as they have been reported to potentially play roles in macro- and microorganism diversity in the environment (29–31). To look for factors that might explain the correlation we observed between sample-site latitude and metagenome KS and AD domain composition, NMDS plots were annotated based on each of these environmental factors (Fig. 3). Collection-site elevations ranged from just above sea level to 928 m above sea level, with many sites falling between 250–500 m above sea level. We found that the higher-elevation sites did not consistently cluster on NMDS plots in an elevation-dependent manner (Fig. 3A). They instead cluster with samples collected at different elevations but in close geographic proximity. Many soil collection sites in the rangeland receive similar amounts of precipitation per year (250–500 mL). Collection sites that received more precipitation occur along the northern and southern coasts of Australia. In neither of the NMDS ordination plots did these geographically distant samples with similarly high precipitation levels cluster together. Instead, they clustered into two geographic groupings that contain other soils from collection sites in close geographic proximity (Fig. 3B). The average daily annual high temperature across our sample collection sites varied from 34.7 °C in the north to 22.7 °C in the south. This continent-wide temperature gradient largely mirrors latitude, as seen in the annotated NMDS ordination plots (Fig. 3C). However, the domain composition at sites where temperatures do not fall into this general continent-wide gradient cluster, once again, based on geographic location.

In analyses of soil-derived 16S gene sequences, the composition of soil bacterial communities has been reported to correlate across long distances with either soil pH or salinity (29–31). Interestingly, a cursory examination of the KS and AD domain NMDS ordination plots annotated according to site pH suggested a similar pH-based correlation with biosynthetic domain composition. In our case, however, this observation appears to be biased by a general latitudinal gradient of pH change across our sample collection sites, with lower-pH sites predominately found in the north and higher-pH sites predominately found in the south. As with temperature, in a number of instances where soil pH falls outside of this general gradient or when soils with different pH occur in close geographic proximity (Fig. 3D, black oval) we see clustering in the NMDS ordination plot by geographic proximity rather than by site pH, suggesting a stronger dependence on latitudinal position than any other factor we measured (Fig. 3E).

While we believe Australia is a model natural laboratory for this type of large-scale ecological study, our results must be viewed in the light of the limitation of this natural system. The very remote nature of northwestern Australia limits sample collection in these areas, and therefore the longitudinal distribution of our samples is more limited than the latitudinal distribution. In addition, with this single-epoch collection approach no consideration is given to either long-term (evolutionary) or short-term (seasonal) effects on changes in environmental microbiome. It will be interesting to extend this analysis to examine more variables, in a larger number of ecologically and geographically diverse samples. When samples

encompassing even more dramatic ecological variation are examined, it is possible that stronger correlations may be observed to some environmental variables. In addition to expanding the geographic diversity of collection sites, it will also be interesting to examine the composition and richness of NPSTs associated with a more diverse collection of common NP classes.

In an effort to explain the latitudinal species richness gradient that is seen for macroorganisms, a variety of ecological, evolutionary, historical, and stochastic models have been proposed (24–27). In ecological models, a global temperature gradient is believed to have been key to the development of the richness gradient as a result of temperature-induced differences in growth and mutation rates at different latitudes (24–27). Our data do not support a temperature-based explanation for the correlation between latitudinal position and biosynthetic domain composition. Causative factors that might explain this correlation could be an as-yet-unrecognized combination of environmental variables we have already measured or as-yet-unknown environmental variables. As with the species richness gradient observed for macroorganisms, which has been studied for over 50 years, the identification of underlying causative factors associated with the biosynthetic domain composition relationships we observed remains elusive and will likely require additional analysis of biosynthetic domain sequences from more geographically and ecologically diverse soils.

Tracking Biomedically Relevant Gene Cluster Families in Australian Soils. One of the ultimate goals of studying NP biosynthetic diversity is to guide the identification of new bioactive metabolites. One way NPST data can assist with this is by identifying environments that are rich in gene cluster families that are known to encode clinically important metabolites (29–35). Biosynthetic gene clusters that encode families of structurally related metabolites generally share a common evolutionary ancestor and therefore exhibit high sequence identity. Using NPST data, this correlation can be exploited to predict the presence of gene cluster families of interest in a metagenome. In this type of analysis, NPSTs are compared using BLAST to a curated database of domain sequences where domains derived from biomedically relevant gene cluster families are appropriately annotated. This BLAST search identifies NPSTs that are more closely related to the biomedically relevant gene cluster families of interest than to any other sequence gene cluster. Due to the common ancestry of individual members within a family of gene clusters, this is a good indicator of a functional relationship between an environmental gene cluster yielding an NPST and a gene cluster family of interest. To facilitate this type of analysis on metagenome sequence data, we have previously developed a web-based sequence analysis platform, eSNaPD (environmental Surveyor of Natural Product Diversity), which automates the identification of metagenomes that are predicted to contain gene cluster families of interest (36, 37). Using eSNaPD we searched our Australian KS and AD NPST datasets for metagenomes predicted to contain gene clusters that are related to NPs of biomedical interest, including NP antibiotics, antiproliferative agents, immunosuppressants, and antifungals. We would propose that sites with the most sequence tags represent the most productive starting points for identifying additional potentially therapeutically improved congeners of NP families of biomedical interest. The locations of the five soils producing the most NPSTs related to six biomedically interesting NPs are shown in Fig. S8. The molecules presented in Fig. S8 were selected because of both their biomedical relevance and the diverse distribution patterns of their NPST data. Equally interesting are NPSTs only distantly related to sequences from functional characterized gene clusters. As the vast majority of bacteria remain uncultured and the vast majority of gene clusters in culture bacteria are silent, most NPST sequence tags found in any environment are not closely related to a functionally characterized gene cluster. The KS and AD domain Chao1 diversity estimates presented in Fig. 1E are therefore reasonable metrics for identifying the ecotypes that are on

CHEMISTRY

MICROBIOLOGY

average the most likely to be productive starting points for potentially identifying novel bioactive NPs. This type of chemical biogeographic data for NP gene cluster families of biomedical interest should prove to be a useful guide for the future discovery of therapeutically relevant NPs from Australian soils.

Conclusion

Secondary metabolism is a critical component of a bacterium's capacity to interact with its surroundings. Although the constitution of the collective bacterial secondary metabolome present in an environment is likely critical to the function of its microbiome, we currently lack an understanding of factors associated with changes in secondary metabolism on a metagenome-wide scale. In an effort to address this deficiency we have coupled targeted sequencing of secondary metabolite biosynthesis domains with metadata collected at geographically distant sites across the continent of Australia. When metadata collected at each soil site were used to look for environmental factors that might correlate with biosynthetic domain richness, the richness appeared to be independent of any sample site characteristics we analyzed and showed little variation across the continent. We did not detect any significant correlations between biosynthetic domain richness patterns and any environmental parameters we measured. Changes in biosynthetic domain composition correlated most closely with changes in collection-site latitude. The other factors we measured, in particular pH, correlated with biosynthetic domain composition differences across subsets of the sites we sampled; however, none of these correlations was as robust as what we observed for latitude. Interestingly, this more closely resembles what has been reported for macroorganisms where nonedaphic factors (i.e., latitude and temperature) can correlate with changes in species composition or richness. For soil microorganisms, changes in edaphic factors (i.e., pH and salinity) have been found to correlate most strongly with changes in species composition or richness (29–31). From a practical perspective, our results indicate that maximizing differences in collection-site latitude may provide a simple means for maximizing the biosynthetic gene cluster diversity used in future novel NP discovery efforts, although a more in-depth analysis of additional environmental variables and combinations of environmental variables from a more geographically diverse collection of soils is needed to be able to understand causative factors driving the differences in biosynthetic composition at different latitudes.

Methods

Around 200 g of topsoil was collected from 397 Australian sites along with a series of environmental variables for each site. Degenerate primers were used to PCR-amplify AD and KS domains before performing DNA sequencing on an Illumina MiSeq platform (602 cycles: 301 × 301). Raw reads from the sequencing were debarcoded and trimmed to remove low-quality reads and bases. Subsequently, OTUs were generated with the Usearch software (23). An NMDS analysis with the Bray–Curtis distance ordination method was performed using the phyloseq R package (28). The resulting OTUs were also queried using eSNaPD for AD and KS NPST annotation against our database gene cluster domains characterized for producing NPs of medical importance (36, 37). A more detailed description of the methods can be found in *SI Methods*.

1. Meinwald J, Eisner T (2008) Chemical ecology in retrospect and prospect. *Proc Natl Acad Sci USA* 105:4539–4540.
2. Rappé MS, Giovannoni SJ (2003) The uncultured microbial majority. *Annu Rev Microbiol* 57:369–394.
3. Rajendhran J, Gunasekaran P (2011) Microbial phylogeny and diversity: Small subunit ribosomal RNA sequence analysis and beyond. *Microbiol Res* 166:99–110.
4. Gilbert JA, Dupont CL (2011) Microbial metagenomics: Beyond the genome. *Annu Rev Mar Sci* 3:347–371.
5. Dewick PM (2002) *Medicinal Natural Products: A Biosynthetic Approach* (Wiley, New York).
6. Banik JJ, Brady SF (2008) Cloning and characterization of new glycopeptide gene clusters found in an environmental DNA megalibrary. *Proc Natl Acad Sci USA* 105:17273–17277.
7. Feng Z, Kallifidas D, Brady SF (2011) Functional analysis of environmental DNA-derived type II polyketide synthases reveals structurally diverse secondary metabolites. *Proc Natl Acad Sci USA* 108:12629–12634.
8. Chang FY, Brady SF (2013) Discovery of indolotryptoline antiproliferative agents by homology-guided metagenomic screening. *Proc Natl Acad Sci USA* 110:2478–2483.
9. Iqbal HA, Feng Z, Brady SF (2012) Biocatalysts and small molecule products from metagenomic studies. *Curr Opin Chem Biol* 16:109–116.
10. Woodhouse JN, Fan L, Brown MV, Thomas T, Neilan BA (2013) Deep sequencing of non-ribosomal peptide synthetases and polyketide synthases from the microbiomes of Australian marine sponges. *ISME J* 7:1842–1851.
11. Gontang EA, Gaudêncio SP, Fenical W, Jensen PR (2010) Sequence-based analysis of secondary-metabolite biosynthesis in marine actinobacteria. *Appl Environ Microbiol* 76:2487–2499.
12. Ziemert N, et al. (2012) The natural product domain seeker NaPDoS: A phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS One* 7:e34064.
13. Lindenmayer D, Burns E, Thurgate N, Lowe A; Commonwealth Scientific and Industrial Research Organization (2014) *Biodiversity and Environmental Change: Monitoring, Challenges, and Direction* (CSIRO Publishing, Clayton, VIC, Australia).
14. Doroghazi JR, et al. (2014) A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nat Chem Biol* 10:963–968.
15. Jensen PR (2010) Linking species concepts to natural product discovery in the post-genomic era. *J Ind Microbiol Biotechnol* 37:219–224.
16. Czárán TL, Hoekstra RF, Pagie L (2002) Chemical warfare between microbes promotes biodiversity. *Proc Natl Acad Sci USA* 99:786–790.
17. Bull AT, Stach JE (2007) Marine actinobacteria: New opportunities for natural product search and discovery. *Trends Microbiol* 15:491–499.
18. Pennisi E (2005) What determines species diversity? *Science* 309:90.
19. Guerin GR, et al. (2017) Opportunities for integrated ecological analysis across inland Australia with standardised data from Ausplots Rangelands. *PLoS One* 12:e0170137.
20. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* 25:1965–1978.
21. Ayuso-Sacido A, Genilloud O (2005) New PCR primers for the screening of NRPS and PKS-I systems in actinomycetes: Detection and distribution of these biosynthetic gene sequences in major taxonomic groups. *Microb Ecol* 49:10–24.
22. Schirmer A, et al. (2005) Metagenomic analysis reveals diverse polyketide synthase gene clusters in microorganisms associated with the marine sponge Discodermia dissoluta. *Appl Environ Microbiol* 71:4840–4849.
23. Edgar RC (2013) UPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* 10:996–998.
24. Pianka ER (1966) Latitudinal gradients in species diversity: A review of concepts. *Am Nat* 100:33–46.
25. Schemske DW, Mittelbach GG (2017) "Latitudinal gradients in species diversity": Reflections on Pianka's 1966 article and a look forward. *Am Nat* 189:599–603.
26. Condamine FL, Sperling FA, Wahlberg N, Rasplus JY, Kergoat GJ (2012) What causes latitudinal gradients in species diversity? Evolutionary processes and ecological constraints on swallowtail biodiversity. *Ecol Lett* 15:267–277.
27. Rohde K (1992) Latitudinal gradients in species diversity: The search for the primary cause. *Oikos* 65:514–527.
28. McMurdie PJ, Holmes S (2013) phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8:e61217.
29. Fierer N, Jackson RB (2006) The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci USA* 103:626–631.
30. Zhalnina K, et al. (2015) Soil pH determines microbial diversity and composition in the park grass experiment. *Microb Ecol* 69:395–406.
31. Lozupone CA, Knight R (2007) Global patterns in bacterial diversity. *Proc Natl Acad Sci USA* 104:11436–11440.
32. Kallifidas D, Brady SF (2012) Reassembly of functionally intact environmental DNA-derived biosynthetic gene clusters. *Methods Enzymol* 517:225–239.
33. Kang HS, Brady SF (2014) Arixanthomycins A-C: Phylogeny-guided discovery of biologically active eDNA-derived pentangular polyphenols. *ACS Chem Biol* 9:1267–1272.
34. Owen JG, et al. (2015) Multiplexed metagenome mining using short DNA sequence tags facilitates targeted discovery of epoxyketone proteasome inhibitors. *Proc Natl Acad Sci USA* 112:4221–4226.
35. Chang FY, Ternei MA, Calle PY, Brady SF (2015) Targeted metagenomics: Finding rare tryptophan dimer natural products in the environment. *J Am Chem Soc* 137:6044–6052.
36. Reddy BV, Milshteyn A, Charlop-Powers Z, Brady SF (2014) eSNaPD: A versatile, web-based bioinformatics platform for surveying and mining natural product biosynthetic diversity from metagenomes. *Chem Biol* 21:1023–1033.
37. Owen JG, et al. (2013) Mapping gene clusters within arrayed metagenomic libraries to expand the structural diversity of biomedically relevant natural products. *Proc Natl Acad Sci USA* 110:11797–11802.
38. White A, et al. (2012) AusPlots Rangelands survey protocols manual (Univ of Adelaide Press, Adelaide, SA, Australia), Version 1.2.9.
39. Tokmakoff A, Sparrow B, Turner D, Lowe A (2016) AusPlots Rangelands field data collection and publication: Infrastructure for ecological monitoring. *Future Gener Comput Syst* 56:537–549.
40. Li H (2016) seqtk: A fast and lightweight tool for processing sequences (Broad Inst., Cambridge, MA).