

ARTICLE

Received 17 Jun 2010 | Accepted 15 Dec 2010 | Published 25 Jan 2011

DOI: 10.1038/ncomms1167

# Ancient origins determine global biogeography of hot and cold desert cyanobacteria

Justin Bahl<sup>1</sup>, Maggie C. Y. Lau<sup>2</sup>, Gavin J. D. Smith<sup>1</sup>, Dhanasekaran Vijaykrishna<sup>1</sup>, S. Craig Cary<sup>3</sup>,  
Donnabella C. Lacap<sup>2</sup>, Charles K. Lee<sup>3</sup>, R. Thane Papke<sup>4</sup>, Kimberley A. Warren-Rhodes<sup>5</sup>, Fiona K. Y. Wong<sup>2</sup>,  
Christopher P. McKay<sup>5</sup> & Stephen B. Pointing<sup>2</sup>

Factors governing large-scale spatio-temporal distribution of microorganisms remain unresolved, yet are pivotal to understanding ecosystem value and function. Molecular genetic analyses have focused on the influence of niche and neutral processes in determining spatial patterns without considering the temporal scale. Here, we use temporal phylogenetic analysis calibrated using microfossil data for a globally sampled desert cyanobacterium, *Chroococcidiopsis*, to investigate spatio-temporal patterns in microbial biogeography and evolution. Multilocus phylogenetic associations were dependent on contemporary climate with no evidence for distance-related patterns. Massively parallel pyrosequencing of environmental samples confirmed that *Chroococcidiopsis* variants were specific to either hot or cold deserts. Temporally scaled phylogenetic analyses showed no evidence of recent inter-regional gene flow, indicating populations have not shared common ancestry since before the formation of modern continents. These results indicate that global distribution of desert cyanobacteria has not resulted from widespread contemporary dispersal but is an ancient evolutionary legacy. This highlights the importance of considering temporal scales in microbial biogeography.

<sup>1</sup> Duke-NUS Graduate Medical School, Singapore 169857. <sup>2</sup> School of Biological Sciences, The University of Hong Kong, Pokfulam Road, Hong Kong, China. <sup>3</sup> Department of Biological Sciences, The University of Waikato, Private Bag 3105, Hamilton, New Zealand. <sup>4</sup> Department of Molecular and Cell Biology, University of Connecticut, Storrs, Connecticut 06269, USA. <sup>5</sup> Space Science Division, NASA Ames Research Center, Moffet Field, California 94035, USA. Correspondence and requests for materials should be addressed to S.B.P. (email: pointing@hku.hk).

Microorganisms occur in all biomes and carry out essential biogeochemical functions and interactions with macroorganisms<sup>1</sup>. Understanding spatio-temporal patterns of microbial diversity and how they are determined is therefore relevant to all realms of ecology, with important implications for understanding ecosystem heterogeneity, conservation value and bioprospecting<sup>2</sup>. Although considerable effort has focused on the environmental drivers of community assembly<sup>3</sup>, relatively little is known about the factors that determine microbial biogeographic patterns for a given taxon, largely due to limitations of ineffective phylogenetic resolution and sampling of taxa and a lack of consideration of temporal scales<sup>4</sup>. Current dogma is supported by key phylogenetic studies on thermophilic prokaryotes, in which a positive correlation between geographic distance and genetic distance for populations has been shown<sup>5,6</sup>. Such studies have addressed spatial variability without considering the temporal scale, and we hypothesized that doing so would yield additional and critical insight on this issue.

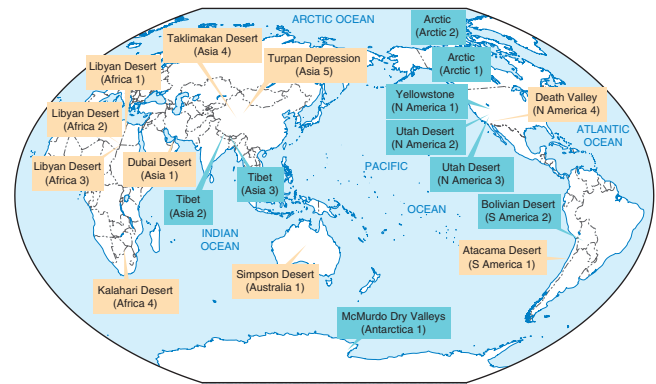
Deserts provide ideal model systems for large-scale microbial biogeographic studies, as they comprise the most abundant terrestrial biome on the Earth<sup>7</sup> and are relatively low energy systems that share readily identifiable major environmental stressors (that is, water availability and temperature<sup>8</sup>) and a relative lack of confounding effects from grazing organisms. Deserts may also be likened to islands, as they are not spatially contiguous. The proximity of hot and cold deserts within the same continental regions provides natural controls to test hypotheses related to global distribution patterns, specifically, the extent that distribution of microorganisms reflects allopatric or selective forces.

Here, we present a comprehensive analysis using a globally derived data set of the hypolithic cyanobacterial genus *Chroococcidiopsis*, which are common primary producers occurring in both hot<sup>9,10</sup> and cold<sup>11,12</sup> extreme arid deserts. A temporally scaled phylogenetic analysis combined with massively parallel pyrosequencing of environmental samples revealed distribution was dependent on contemporary climate and *Chroococcidiopsis* variants were specific to either hot or cold deserts. Within each climatically defined group, strong endemic signals were also recorded. Temporal phylogenies showed no evidence of recent inter-regional gene flow and indicated populations have not shared common ancestry since before the formation of modern continents. No evidence for distance-related patterns was detected. This study emphasizes the value of considering temporal scales in addition to spatial and environmental factors in microbial biogeography.

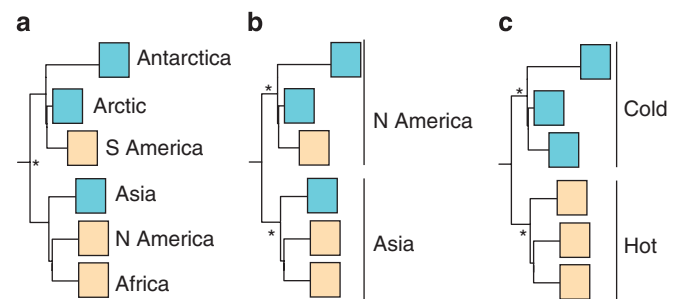
## Results

**Global sampling of desert environments.** We assembled an extensive sequence-based data set for the ubiquitously distributed desert cyanobacterial genus *Chroococcidiopsis* from every major desert on the Earth (Fig. 1). These habitats included both hot and cold desert climates. We successfully recovered environmental DNA from hypolithic cyanobacterial biomass on translucent quartz. Clone libraries generated from each environmental sample were screened for *Chroococcidiopsis* phylotypes by analysis of 16S rRNA gene sequences and their 16S rRNA, 5.8S ITS and 23S rRNA loci were then sequenced for phylogenetic analysis. A total of 73 unique phylotypes were identified and sequenced (Supplementary Table S1).

**Hypothetical scenarios for microbial biogeography.** This globally derived data set allowed for the testing of specific ecological and evolutionary scenarios (Fig. 2). If microbial dispersal were ubiquitous without environmental selection, random patterns of distribution would emerge (Fig. 2a). These deserts would effectively present a single-microbial habitat and biogeographic province<sup>3</sup>. Alternatively, if geographic isolation, rather than environmental selection, were



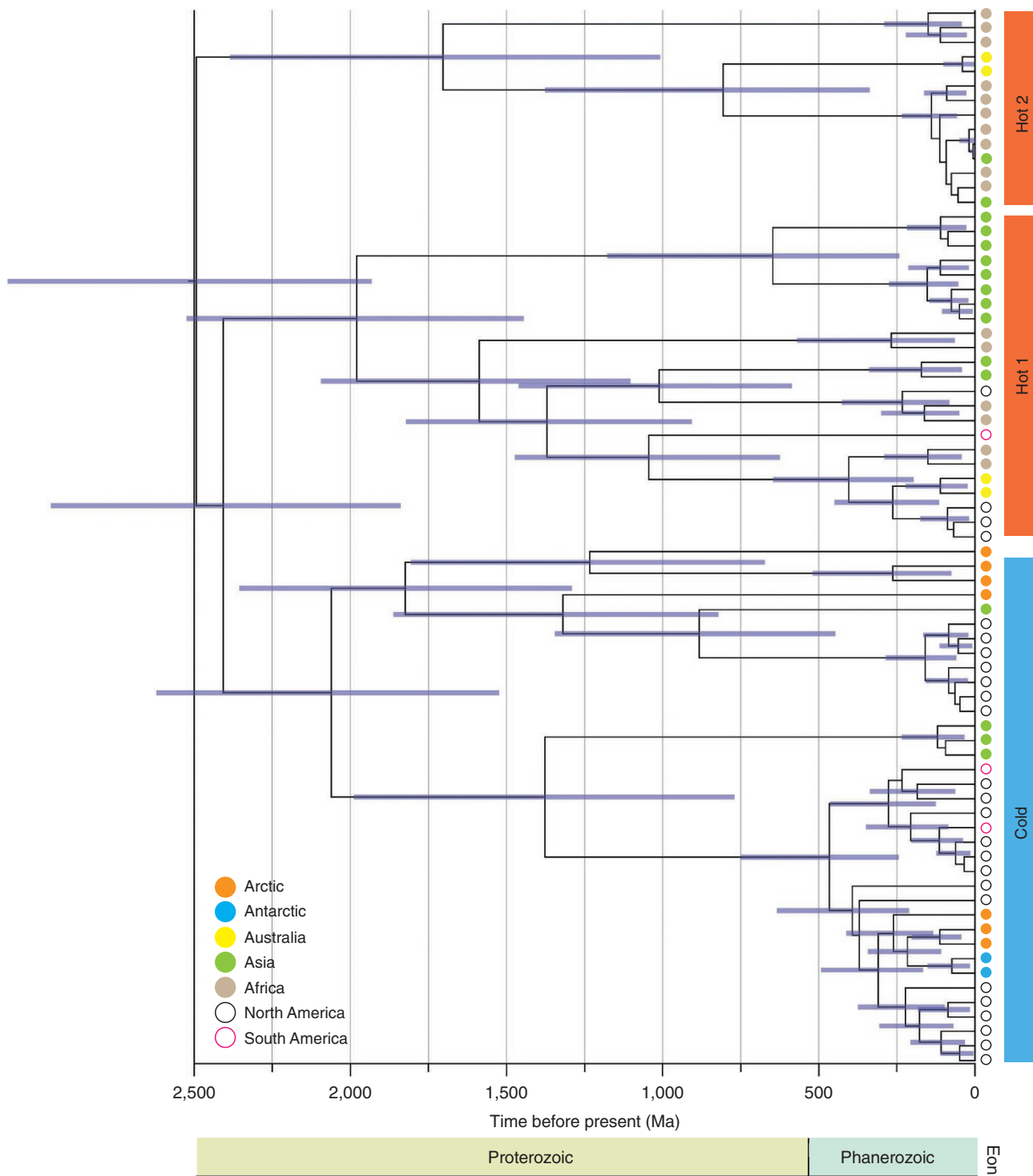
**Figure 1 | Global distribution of *Chroococcidiopsis* variants.** The map indicates hot (brown boxes) and cold (blue boxes) desert locations from which variants were recovered. Detailed site descriptions and climatic information are given in Supplementary Table S1.



**Figure 2 | Idealized phylogenies of hypothetical scenarios for global distribution of desert cyanobacteria.** Scenario **a**, assumes ubiquitous distribution resulting in mixed geographic regions and mixed environments within a phylogeny. Here, root divergence is a relatively recent event. Scenario **b**, assumes allopatric speciation resulting in distinct geographic regions and mixed environments within a phylogeny (that is, N America and Asia are monophyletic groups). Here, root divergence times correspond with formation of continents. Scenario **c**, assumes environmental selection corresponding to global climatic change, resulting in a phylogeny with mixed geographic regions and distinct environments. Here, the root divergence times are significantly older than known ages of formation of continents. Asterisks indicate monophyletic groups. Brown boxes and blue boxes indicate hot and cold desert locations, respectively.

the main driver of diversity then location-specific lineages would arise in different provinces regardless of microhabitat (Fig. 2b). For example, hot and cold deserts in the North America would share common ancestry. Finally, in a scenario in which environmental selection drives diversification lineages consisting of bacteria from cold and hot habitats would be observed (Fig. 2c). For example, the cyanobacteria from hot desert in Death Valley would share ancestry with hot deserts from other continents rather than with cyanobacteria from cold desert in Utah.

**Ancient times of common ancestry for extant *Chroococcidiopsis*.** A temporal phylogeny calibrated using fossil records<sup>13,14</sup> that have been shown as robust calibration points for cyanobacterial phylogenies<sup>15</sup>, was constructed using a Bayesian relaxed-clock analysis of the complete, unambiguously aligned, 16S-ITS-23S rRNA gene regions to estimate the divergence times of the globally distributed *Chroococcidiopsis* variants (Fig. 3, Supplementary Figs S1–S5 show all individual and combined gene trees). We used a relaxed phylogenetic approach<sup>16</sup> to estimate phylogeny and divergence times while



**Figure 3 | Temporal phylogeny for *Chroococciopsis* variants.** Variants were recovered from hot and cold deserts worldwide as shown in Supplementary Table S1. The tree was generated using a Bayesian relaxed-clock phylogenetic approach<sup>16</sup> of the 16S-ITS-23S rDNA regions to estimate divergence dates. The age of the common ancestor for *Chroococciopsis* was estimated with a 95% Bayesian confidence interval, using fossil ancestors for calibration<sup>13,14</sup>. Blue bars at nodes indicate 95% credible intervals for divergence events. Temporal scale is shown in millions of years.

taking into account uncertainties in evolutionary rates and calibration times<sup>17</sup>. Briefly, a 16S rRNA gene cyanobacterial phylogeny was constructed using the age of two fossilized ancestors as calibration points for estimation of evolutionary rates and timing of divergence events (Supplementary Fig. S1). The estimated age of the common ancestor for *Chroococciopsis* was then incorporated as a

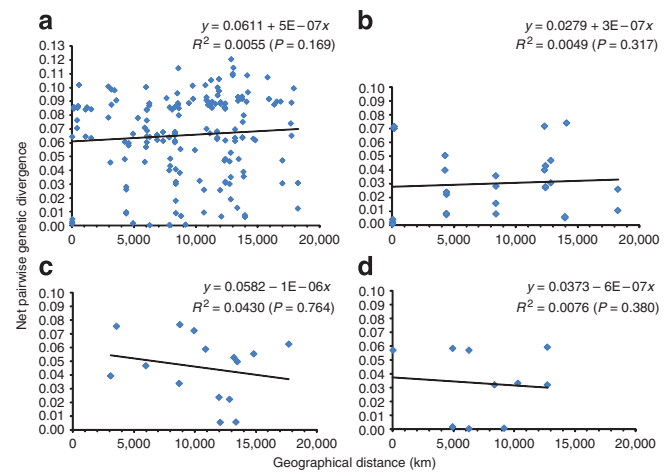
previous constraint on subsequent Bayesian relaxed-clock analysis of the complete, unambiguously aligned, 16S-ITS-23S rRNA gene regions to estimate the divergence times of the globally distributed *Chroococciopsis* variants (Supplementary Figs S2–S5).

Our results showed hot and cold desert variants were evolutionarily distinct. All cold variants were monophyletic, whereas

hot variants formed two groups with one sharing an ancient common ancestor with cold variants  $\sim 2.4$  Ga (Fig. 3, hot clade 1) and another ancestral to all extant *Chroococcidiopsis* variants ( $\sim 2.5$  Ga; Fig. 3, hot clade 2). The temporal phylogeny suggests that the time of most recent common ancestry to all contemporary variants was 2.5 Ga (range: 3.1–1.9 Ga) with a mean substitution rate estimate of  $2.7 \times 10^{-5}$  substitutions/site/Ma (95% Bayesian confidence interval 2.0– $3.5 \times 10^{-5}$ ; Fig. 3, Supplementary Table S2). Within each clade, regionally distinct populations were apparent. For example, within the hot clade all variants from the Turpan Depression (Asia 5 in Supplementary Figs S1–S5) were monophyletic and share a common ancestor from  $\sim 600$  Ma. Region-specific lineages were apparent towards the tips of the phylogenetic trees, but the most recent times of common ancestry were  $\sim 100$  Ma (Fig. 3). Evidence of gene flow between climatically similar deserts was also supported from the phylogenetic analysis. For example, within the last 100 million years a population of African variants from the Libyan Desert, Egypt (Africa 2) likely established a founder population in the Taklimakan Desert, China (Asia 4) and links African and Asian variants within the hot 2 clade (Supplementary Fig. S2). But, in all cases, where such events could be detected these were ancient divergences and rare.

**Absence of invasive colonization between deserts.** We also carried out massively parallel rRNA gene directed metagenomic 454 pyrosequencing of hot and cold desert environmental samples, targeting the highly variable V5–V7 16S rRNA gene region to identify possible ‘rare’ phylotypes not recovered in our clone libraries<sup>18,19</sup> (Supplementary Table S3). This analysis showed that contemporary hot and cold clade *Chroococcidiopsis* variants were exclusive to hot and cold deserts, respectively, with the exception of a single cold clade variant that was recovered from the hot desert at very low abundance ( $< 0.05\%$ ) within the assemblage. Hot variants did, however, share greatest phylogenetic affiliation with variants from their own rather than any other climatically similar desert. Therefore, dispersal and establishment of hot variants in cold deserts ( $P=0$ ), and vice versa ( $P=0.002$ ), were extremely rare events.

**Distance–decay patterns do not explain global scale distribution.** Effects of climatic and substrate variables were evaluated by analysis of molecular variance (AMOVA, incorporating F Statistic) and phylogenetic test (P-test), whereas the effects of distance-related variables were assessed using the Mantel Statistic. Contemporary climate classification based on long-term mean annual temperature and precipitation data was significant in differentiating hot and cold clades (AMOVA,  $F_{CT}=0.29147$ ,  $P<0.00001$ ). The phylogenetically informed delineation into ‘hot’ and ‘cold’ groupings provided the most parsimonious explanation (P-test,  $P<0.001$ ) for the observed tree topology. On a global scale, genetic differentiation was not related to geographic distance (Mantel,  $R^2=0.0055$ ,  $P=0.169$ ; Fig. 4a) or altitude (Mantel,  $R^2=0.0011$ ,  $P=0.601$ ) for the homogenous substrate either within or between these major clades. We further tested the genetic and geographic distance relationship of the phylogenetically identified cold, hot 1 and hot 2 clades by the same method. On this more refined genetic scale, we were also unable to infer a correlation between genetic divergence and geographic distance (Fig. 4b–d). Analysis of relatively close desert locations showed a lack of relatedness between variants from the hot Death Valley desert versus the cold Utah Desert in North America (AMOVA,  $F_{ST}=0.84053$ ,  $P=0.00586$ ), the warm Turpan Depression versus the cold Tibetan tundra in China (AMOVA,  $F_{ST}=0.74756$ ,  $P=0.00391$ ), and the warm Atacama Desert versus the cold high altitude Bolivian Desert in South America (AMOVA,  $F_{ST}=0.93831$ ,  $P<0.00001$ ), whereas Antarctic variants were most closely affiliated with Arctic variants ( $F_{ST}=0.35368$ ,  $P=0.18359$ ).



**Figure 4 | Relationship between genetic divergence and geographic distance among *Chroococcidiopsis* variants.** Genetic differentiation was not significantly related to geographic distance (a) on the global scale ( $N=19$ ,  $n=171$ ), or for each of the phylogenetically defined clusters (b) cold ( $N=8$ ,  $n=28$ ), (c) hot 1 ( $N=6$ ,  $n=15$ ) or (d) hot 2 ( $N=5$ ,  $n=10$ ); where  $N$  denotes number of samples and  $n$  the resultant number of pairwise comparisons. The best-fit linear regression function, the coefficient of determination ( $R$ -square) and significance ( $p$ ) of Mantel test are displayed for individual regression plots. A significance level ( $\alpha$ ) of 0.05 was applied.

## Discussion

The temporal phylogeny generated divergence time estimates that broadly concur with estimates for the onset of widespread aridity on the Earth ( $\sim 1.8$  Ga)<sup>8</sup>. Interestingly, the estimated substitution rate was approximately an order of magnitude lower than estimates for other bacterial groups<sup>20</sup>. This further supports the suggestion that a universal evolutionary rate may not exist for all bacteria<sup>20</sup>. The majority of divergences within the temporal phylogeny that explain diversity of extant *Chroococcidiopsis* variants were ancient and pre-date the estimated onset of contemporary aridity for many locations. For example the oldest contemporary arid region is thought to be the Namib Desert ( $\sim 80$  Ma<sup>21</sup>). The North African, the Polar, the South American and the Tibetan aridity have likely persisted since 30–45 Ma<sup>22–27</sup>, whereas others such as the Death Valley, the Simpson and the Taklimakan deserts may be relatively recent<sup>28–30</sup>. Despite regionally specific monophyletic clades within both hot and cold clades, the times of common ancestry for *Chroococcidiopsis* variants pre-dated estimates for contemporary aridity in these desert regions. This suggests that the global distribution for *Chroococcidiopsis* has been limited by barriers to long distance dispersal and/or invasive colonization, with regional gene pools maintained over geological timescales. Environmental selection may also have exerted a major role in colony establishment in different geographical regions. Similar to our results, DNA fingerprinting studies of *Pseudomonas* isolates from soil showed that strains were not globally mixed, but rather regionally endemic strains had evolved<sup>31</sup>.

The high-coverage sequencing of environmental samples further indicated that the hot and cold delineation for *Chroococcidiopsis* variants is valid ecologically, as variants were exclusive to either hot or cold deserts. From our data it is reasonably concluded that dispersal and establishment of hot variants in cold deserts, and vice versa, are extremely rare events. As phylotypes were more similar to those from their own rather than any other climatically similar desert, this also suggests a lack of dispersal between similar habitats. The temporal phylogeny therefore strongly indicates that variants within hot and cold clades do not readily disperse and colonize climatically similar deserts. Although founder populations may

establish colonies in other climatically similar regions, our data suggests these events were rare. Whereas it is possible that a population of African variants from the Libyan Desert, Egypt, established a founder population in the Taklimakan Desert, China, within the last 100 million years, linking both African and Asian variants within the hot 2 clade, similar events were apparently rare despite the very long evolutionary timescales. Additional examples were evident in the phylogenetic trees, but it was unlikely that any establishment of new populations in these geographically distant locales was a recent event. This further indicates that dispersal events between these desert locations could be limiting and/or invasive colonization rarely succeeds. Our phylogeny thus becomes predictive for evolutionary lineage in a given climatically defined desert.

The different populations recovered from locations in relatively close proximity, but experiencing different climates illustrated how differentiation of *Chroococcidiopsis* variants can be explained by adaptation to climate-driven selection pressures on the bacterial population rather than distance between populations. For example, the lack of relatedness between variants from the hot Death Valley desert versus the cold Utah Desert in North America, the warm Turpan Depression versus the cold Tibetan desert in China or the warm Atacama Desert versus the cold high altitude Bolivian Desert in South America. This contrasted with highly similar populations from Antarctic and Arctic locations, strongly suggesting geographic distance has little effect on divergence or relatedness between populations.

We also demonstrated that distance-related patterns were not apparent on a global scale for variants adapted to similar environments (that is, within hot or cold clades), suggesting that distance-decay relationships may not be applicable to this globally distributed bacterium. This is in contrast to the notion of distance-decay relationships between genetic and geographic distance among microbial populations<sup>2–6</sup>. Recent experimental studies have demonstrated that at the community level, distance-decay patterns may be observed in the short term (<7 days) but have a relatively minor role over longer time periods<sup>32</sup>. Although an increased interest in microbial biogeography has led to a re-examination of the distance-decay explanation of bacterial distribution patterns for various communities<sup>2,3,4</sup>, there nonetheless remains insufficient data to reject this hypothesis or support alternatives to describe microbial distribution, community structure or genetic variation.

We hypothesize that acquisition of adaptive traits facilitating speciation into cold desert variants was an ancient event that evolved in globally dispersed ancestral hot-adapted *Chroococcidiopsis* variants. The ancestral variant is a likely candidate for early terrestrial life during the Earth's long arid history<sup>8,33</sup>. We demonstrated that although a desert cyanobacterium such as *Chroococcidiopsis* may notionally possess the characteristics for ubiquitous dispersal due to aeolian transport of desert particulates and its desiccation/radiation tolerance<sup>34,35</sup>, this is not reflected in contemporary colonization patterns. The existence of regionally isolated gene pools of hot and cold *Chroococcidiopsis* variants strongly supports the concept that widespread contemporary dispersal is not common, and that relationships reflect ancient historical legacies<sup>36</sup>. Although strong selection for hot and cold variants has occurred, this pre-dates contemporary climatic selective pressures and so indicates that the complex interaction of allopatric and selective forces can be more fully understood by considering temporal scales in microbial biogeography. This also highlights a value for local microbial diversity and suggests that microorganisms are likely to face threats of reduced biodiversity due to local extinction in a manner similar to macroorganisms. Temporally calibrated phylogenetic data sets such as these may in future also help to inform issues in the reconstruction of ancient geomorphology and climate, in a manner analogous to records of colonization by macroorganisms<sup>37</sup>.

## Methods

**Environmental samples.** We targeted the hypolithic niche (cyanobacterial colonization on the ventral surface of translucent quartz rock), as this inert and relatively homogeneous substrate is ubiquitous in desert pavement terrain. Hypoliths are photoautotrophs and so localized substrate-related effects on communities are minimized. Hypolithic communities dominated by cyanobacteria are clearly distinct from the surrounding soil biota<sup>12</sup> and long lived<sup>10</sup>. The combination of stable environment and a long-lived community reduces potential for bias due to short-term temporal variability. The cyanobacterial genus *Chroococcidiopsis* has monophyletic rRNA gene defined origins<sup>38</sup> and cyanobacteria have a documented micro-fossil<sup>13,14</sup> record that allows temporal phylogenies to be calibrated.

We collected colonized white quartz rocks (hypoliths,  $N=54$ ) from desert pavement terrain in desert locations worldwide (Fig. 1, Supplementary Table S1). Geographical distance between sites was calculated using the law of haversines. At least three hypoliths were collected from each location. Biomass was recovered aseptically and environmental DNA purified by phenol–chloroform extraction and ethanol precipitation. For each location we recorded precise location and this was mapped via geographic information system to a global climate map based on long-term mean annual temperature and rainfall<sup>7,8</sup>. The quartz (SiO<sub>2</sub>) substrate was characterized using standard USGS mineralogical diagnostic criteria (<http://www.usgs.gov>).

**Recovery of *Chroococcidiopsis* variants.** Clone libraries (colonies/library,  $N\leq 50$ ; total  $N\leq 950$ ) were generated from hypolith samples from each location using cyanobacteria-specific primers (359F-2763R)<sup>39,40</sup> spanning a contiguous region of loci with incremental rates of evolution including 16S rDNA, ITS and 23S rDNA. Libraries were screened for *Chroococcidiopsis* variants by sequencing (359F-781R)<sup>40</sup> (BigDye Terminator 3.1, 3730 Genetic Analyzer, Applied Biosystems). Individual phylotypes were identified from the environmental sample clone libraries by phylogenetic analysis of 16S rRNA gene sequences (~550 bp) using distance criteria optimized with the Generalized Time Reversible model (GTR + I + G) (PAUP v4.0 (ref. 41)). All unique clones that affiliated phylogenetically with *Chroococcidiopsis* were then sequenced along with the entire 16S rRNA, 5.8S ITS and 23S rRNA gene region comprising ~4,300 bp (Supplementary Table S4). Ambiguously aligned regions were excluded from all analyses. The final alignment was 3,645 bp ( $N=73$ ). Model selection was achieved using ModelTest<sup>42</sup> using the Akaike Information Criterion to choose the best-fit model. In all cases, the best-fit model was the GTR + I + G model and this was applied to each analysis.

**Relaxed-clock phylogenetic analyses.** Temporal phylogenies were generated using a Bayesian 'relaxed molecular clock' approach with a Bayesian skyride coalescent prior<sup>16,43</sup>. This prior does not require strong *a priori* decisions about the demographic history of the population (that is, number of coalescent events) and was used in all analyses. Initially, 16S rRNA genes of a broad cross-section of cyanobacterial orders, including Chroococcales, Nostocales, Oscillatoriales, Pleurocapsales, Prochlorales, Stigonematales and plastids available from the NCBI GenBank were used ( $N=224$ , Supplementary Fig. S1). Two fossilized ancestors for the cyanobacterial families were used as calibration points for estimation of evolutionary rates and timing of divergence events<sup>13,14</sup>. Tests for congruence between the complete data set (Supplementary Fig. S2) and among trees generated from different loci (Supplementary Figs S3–S5) were made using the Shimodaira–Hasegawa test<sup>44</sup> (Supplementary Table S5). The estimated age of the common ancestor for *Chroococcidiopsis* was then incorporated as a previous constraint on subsequent Bayesian relaxed-clock analysis of the complete, unambiguously aligned, 16S-ITS-23S rRNA gene regions (~3.6 kbp) to estimate the divergence times of the globally distributed *Chroococcidiopsis* variants (Supplementary Fig. S2).

**Massively parallel rRNA gene-directed pyrosequencing.** Environmental DNA from representative hot (Turpan Depression) and cold (central Tibet) desert hypoliths within the same continental region were selected for massively parallel pyrosequencing. Amplicons of the hyper-variable V5–V7 region of the 16S rRNA gene<sup>18</sup> were pyrosequenced using the Roche 454 Life Sciences GS-FLX sequencer (Taxon Biosciences). Pyrosequencing flowgrams were filtered and de-noised using a variation of the AmpliconNoise suite designed for FLX Titanium sequences. In short, reads with at least one flow of signal intensity between 0.5 and 0.7 or a cycle of four nucleotide flows (ATGC) that failed to give a signal >0.5 (both signs of noise) before flow no. 400 were discarded, and all reads were truncated at flow no. 600. Pyrosequencing noise was removed from filtered flowgrams using the PyroNoise<sup>19</sup> program with default settings, and PCR noise was removed from the resulting sequences using the SeqNoise<sup>19</sup> program ( $\sigma=0.033$ ,  $cs=0.08$ ) after they were truncated at 400 bp to further reduce noise. Last, the sequences were checked for PCR chimeras using Perseus ( $\alpha=-7.5$ ,  $\beta=0.5$ ). The output sequences were annotated using Metagenome Rapid Annotation using Subsystem Technology to identify cyanobacteria-affiliated sequences<sup>45</sup>. Sequences that were 97% homologous to *Chroococcidiopsis* were further confirmed by DOTUR<sup>46</sup> and online BLAST tool (Supplementary Table S3).

**Nucleotide sequence accession.** Nucleotide sequences generated in this study have been deposited to the NCBI GenBank database under accession numbers FJ805842–FJ805957 (temporal phylogeny), HM453992–HM454160 (cold desert pyrosequencing) and HM489009–HM489864 (hot desert pyrosequencing).

**Statistical analyses.** Effects of climatic and substrate variables were evaluated by AMOVA (incorporating  $F_{ST}$ ) on nucleotide sequences and P-test<sup>47</sup> on tree topologies, whereas the effects of distance related variables (location, altitude) were assessed using the Mantel Statistic (Arlequin v3.1 (ref. 48)) with significance testing against 1,000 permutations. Net genetic divergences were calculated using MEGA v4.1 (ref. 49). Normality was ensured before the computation of power functions using Linear Least Squares Fit regression analysis.

## References

- Bell, T., Newman, J. A., Silverman, B. W., Turner, S. L. & Lilley, A. K. The contribution of species richness and composition to bacterial services. *Nature* **436**, 1157–1160 (2005).
- Prosser, J. I. *et al.* The role of ecological theory in microbial ecology. *Nat. Rev. Microbiol.* **5**, 384–392 (2007).
- Hughes-Martiny, J. B. *et al.* Microbial biogeography: putting microorganisms on the map. *Nat. Rev. Microbiol.* **4**, 102–112 (2006).
- Green, J. & Bohannan, B. J. M. Spatial scaling of microbial biodiversity. *Trends Ecol. Evol.* **21**, 501–507 (2006).
- Papke, R. T., Ramsing, N. B., Bateson, M. M. & Ward, D. M. Geographical isolation in hot spring cyanobacteria. *Environ. Microbiol.* **5**, 650–659 (2003).
- Whitaker, R. J., Grogan, D. W. & Taylor, J. W. Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* **301**, 976–978 (2003).
- Peel, M. C., Finlayson, B. L. & McMahon, T. A. Updated world map of the Koppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.* **11**, 1633–1644 (2007).
- Thomas, D. S. G. (ed.) In *Arid Zone Geomorphology* 2nd edition, 3–12 (Wiley, 1997).
- Schlesinger, W. H. *et al.* Community composition and photosynthesis by photoautotrophs under quartz pebbles, southern Mojave Desert. *Ecology* **84**, 3222–3231 (2003).
- Warren-Rhodes, K. A. *et al.* Hypolithic bacteria, dry limit of photosynthesis and microbial ecology in the hyperarid Atacama Desert. *Microbiol. Ecol.* **52**, 389–398 (2006).
- Cockell, C. S. & Stokes, M. D. Widespread colonization by polar hypoliths. *Nature* **431**, 414–415 (2004).
- Pointing, S. B. *et al.* Highly specialized microbial diversity in hyper-arid polar desert. *Proc. Natl Acad. Sci. USA* **106**, 19964–19969 (2009).
- Tomitani, A., Knoll, A. H., Cavanaugh, C. M. & Ohno, T. The evolutionary diversification of cyanobacteria: molecular phylogenetic and paleontological perspectives. *Proc. Natl Acad. Sci. USA* **103**, 5442–5447 (2006).
- Schopf, J. W. & Packer, B. M. Early Archean (3.3-billion to 3.5-billion-year-old) microfossils from Warrawoona Group, Australia. *Science* **237**, 70–73 (1987).
- Falcon, L. I., Magallon, S. & Castillo, A. Dating the cyanobacterial ancestor of the chloroplast. *ISME J.* **4**, 777–783 (2010).
- Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).
- Drummond, A. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88 (2006).
- Ashby, M. N., Rini, J., Mongodin, E. F., Nelson, K. E. & Dimster-Dirk, D. Serial analysis of rRNA genes and the unexpected dominance of rare members of microbial communities. *Appl. Environ. Microbiol.* **73**, 4532–4542 (2007).
- Quince, C. *et al.* Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat. Methods* **6**, 639–641 (2009).
- Kou, C.-H. & Ochman, H. Inferring clocks when lacking rocks: the variable rates of molecular evolution in bacteria. *Biol. Direct.* **4**, 35 (2009).
- Ward, J. D., Steely, M. K. & Lancaster, N. On the antiquity of the Namib. *S. Afr. J. Sci.* **79**, 175–183 (1983).
- Sarnthein, M. In *Initial Reports of the Deep Sea Drilling Project 41* (eds Lancelot, Y. & Seibold, E.) 939–959 (US Government Printing Office, 1978).
- Rowley, D. B. & Currie, B. S. Paleo-altimetry of the late Eocene to Miocene Lunpola Basin, central Tibet. *Nature* **439**, 677 (2006).
- Dunai, T. J. *et al.* Oligocene-Miocene age of aridity in the Atacama Desert revealed by exposure dating of erosion-sensitive landforms. *Geology* **33**, 321–324 (2005).
- Moran, K. *et al.* The Cenozoic paleoenvironment of the Arctic Ocean. *Nature* **441**, 601–605 (2006).
- Barrett, P. J. Antarctic paleoenvironments through Cenozoic time. *Global. Ecol. Biogeog.* **15**, 121–142 (1996).
- Livermore, R., Hillebrand, C.-D., Meredith, M. & Eagles, G. Drake Passage and Cenozoic climate: an open and shut case? *Geochem. Geophys. Geosystems* **8**, Q01005 (2007).
- Sun, J. & Liu, T. The age of the Taklimakan Desert. *Science* **312**, 1621 (2006).
- Bowler, J. M. Aridity in Australia: origins and expression in aeolian landforms and sediments. *Earth Sci. Rev.* **12**, 279–310 (1976).
- Lowenstein, T. K. *et al.* 200k.y. paleoclimate record from Death Valley salt core. *Geology* **27**, 3–6 (1999).
- Cho, J.-C. & Tiedje, J. M. Biogeography and degree of endemicity of fluorescent *Pseudomonas* strains in soil. *Appl. Environ. Microbiol.* **66**, 5448–5456 (2000).
- Bell, T. Experimental tests of the bacterial distance-decay relationship. *ISME J.* **4**, 1357–1365 (2010).
- Wynn-Williams, D. D. In *The Ecology of Cyanobacteria* 1st edn (eds Whitton, B. A. & Potts, M.) 341–366 (Kluwer, 2000).
- Billi, D., Friedmann, E. I., Hofer, K. G., Caiola, M. G. & Ocampo-Friedmann, R. Ionizing-radiation resistance in the desiccation-tolerant cyanobacterium *Chroococcidiopsis*. *Appl. Environ. Microbiol.* **66**, 1489–1492 (2000).
- Kellog, C. A. & Griffin, D. W. Aerobiology and the global transport of desert dust. *Trends Ecol. Evol.* **21**, 638–644 (2006).
- Takacs-Vesbach, C., Mitchell, K., Jackson-Weaver, O. & Reysenbach, A.-L. Volcanic calderas delineate biogeographic provinces among Yellowstone thermophiles. *Environ. Microbiol.* **10**, 1681–1689 (2008).
- Convey, P. *et al.* Antarctic terrestrial life—challenging the history of the frozen continent. *Biol. Rev.* **83**, 103–117 (2008).
- Fewer, D., Friedl, T. & Budel, B. *Chroococcidiopsis* and heterocyst-differentiating cyanobacteria are each other's closest living relatives. *Mol. Phylogenet. Evol.* **23**, 82–90 (2002).
- Lane, D. *Nucleic Acid Techniques in Bacterial Systematics* (eds Stackebrandt, E. & Goodfellow, M.) 115–117 (John Wiley & Sons, 1991).
- Nubel, U., Garcia-Pichel, F. & Muyzer, G. PCR primers to amplify 16S rRNA genes from cyanobacteria. *Appl. Environ. Microbiol.* **63**, 3327–3332 (1997).
- Swofford, D. L. *PAUP\**. *Phylogenetic Analysis Using Parsimony (\*and Other Methods)* Version 4 (Sinauer Associates, 2000).
- Posada, D. & Crandall, K. A. Model test: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818 (1998).
- Minin, V. N., Bloomquist, E. W. & Suchard, M. A. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol. Biol. Evol.* **25**, 1459–1471 (2008).
- Shimodaira, H. & Hasegawa, M. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16**, 1114–1116 (1999).
- Meyer, F. *et al.* The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**, 386 (2008).
- Schloss, P. D. & Handelsman, J. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl. Environ. Microbiol.* **71**, 1501–1506 (2005).
- Maddison, W. P. & Maddison, D. R. Mesquite: a modular system for evolutionary analysis Version 2.0. <http://mesquiteproject.org> (2007).
- Excoffier, L., Laval, G. & Schneider, S. Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evol. Bioinforma. Online* **1**, 47–50 (2005).
- Tamura, K., Dudley, J., Nei, M. & Kumar, S. *MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0*. *Mol. Biol. Evol.* **24**, 1596–1599 (2007).

## Acknowledgments

We thank the following for critical discussion on earlier versions of this work: Yuki Chan, Don Cowan, Alfonso Davilla, Alex Heri, Wayne Pollard, Mark Stevens. This research was supported by the NASA Astrobiology Science and Technology for Exploring Planets (ASTEP) Programme and the Hong Kong Research Grants Council (Grant numbers HKU7733/08 HKU7763/10). J.B., G.J.D.S. and D.V. are supported by the Duke-NUS Signature Research Program funded by the Agency for Science, Technology and Research, and the Ministry of Health, Singapore.

## Author contributions

J.B., M.C.Y.L., G.J.D.S. and D.V. contributed equally to this work. C.P.M. and S.B.P. conceived the study; M.C.Y.L., C.P.M., S.B.P. and K.A.W.-R. collected samples; D.C.L., M.C.Y.L. and F.K.Y.W. conducted Sanger sequencing and phylogenetic analyses; C.K.L. and M.C.Y.L. analysed pyrosequencing data; J.B., M.C.Y.L. and D.V. performed evolutionary analyses; M.C.Y.L. carried out statistical analyses; J.B., S.C.C., M.C.Y.L., C.P.M., R.T.P., S.B.P., G.J.D.S. and D.V. interpreted the results; J.B., C.P.M., S.B.P., G.J.D.S. and D.V. wrote the paper; all authors discussed the results and commented on the manuscript.

## Additional information

**Accession codes.** Nucleotide sequences have been deposited in Genbank under accession codes FJ805842-FJ805957, HM453992-HM454160 and HM489009-HM489864.

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Bahl, J. *et al.* Ancient origins determine global biogeography of hot and cold desert cyanobacteria. *Nat. Commun.* 2:163 doi: 10.1038/ncomms1167 (2011).

**License:** This work is licensed under a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/>