npg

ORIGINAL ARTICLE

Stone-dwelling actinobacteria *Blastococcus saxobsidens*, *Modestobacter marinus* and *Geodermatophilus obscurus* proteogenomes

Haïtham Sghaier[1], Karima Hezbri[2], Faten Ghodhbane-Gtari[2], Petar Pujic[3], Arnab Sen[4], Daniele Daffonchio[5,6], Abdellatif Boudabous[2], Louis S Tisa[7], Hans-Peter Klenk[8], Jean Armengaud[9], Philippe Normand[3] and Maher Gtari[2]

[1]National Center for Nuclear Sciences and Technology, Sidi Thabet Technopark, Ariana, Tunisia; [2]Laboratoire Microorganismes et Biomolécules Actives, Université de Tunis Elmanar (FST) and Université de Carthage (INSAT), Tunis, Tunisia; [3]Université Lyon 1, Université de Lyon, CNRS, Ecologie Microbienne, UMR 5557, Villeurbannes, France; [4]NBU Bioinformatics Facility, Department of Botany, University of North Bengal, Siliguri, India; [5]BESE, Biological and Environmental Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Kingdom of Saudi Arabia; [6]Department of Food, Environmental and Nutritional Sciences (DeFENS), University of Milan, Milan, Italy; [7]Department of Molecular, Cellular and Biomedical Sciences, University of New Hampshire, Durham, NH, USA; [8]School of Biology, Newcastle University, Newcastle upon Tyne, UK and [9]CEA, DSV, IBITEC-S, Lab Innovative Technologies for Detection and Diagnostic, Bagnols-sur-Cèze, France

The Geodermatophilaceae are unique model systems to study the ability to thrive on or within stones and their proteogenomes (referring to the whole protein arsenal encoded by the genome) could provide important insight into their adaptation mechanisms. Here we report the detailed comparative genome analysis of *Blastococcus saxobsidens* (Bs), *Modestobacter marinus* (Mm) and *Geodermatophilus obscurus* (Go) isolated respectively from the interior and the surface of calcarenite stones and from desert sandy soils. The genome-scale analysis of Bs, Mm and Go illustrates how adaptation to these niches can be achieved through various strategies including 'molecular tinkering/opportunism' as shown by the high proportion of lost, duplicated or horizontally transferred genes and ORFans. Using high-throughput discovery proteomics, the three proteomes under unstressed conditions were analyzed, highlighting the most abundant biomarkers and the main protein factors. Proteomic data corroborated previously demonstrated stone-related ecological distribution. For instance, these data showed starvation-inducible, biofilm-related and DNA-protection proteins as signatures of the microbes associated with the interior, surface and outside of stones, respectively.
The ISME Journal (2016) 10, 21–29; doi:10.1038/ismej.2015.108; published online 30 June 2015

Introduction

Geodermatophilaceae are an actinobacterial family (Normand, 2006) in the order Geodermatophiliales (Sen *et al.*, 2014) that comprises three genera: *Geodermatophilus*, *Blastococcus* and *Modestobacter* initially isolated from desert soils (Luedemann, 1968), sea water (Ahrens and Moll, 1970) and Antarctic regolith (Mevs *et al.*, 2000), respectively. These three genera have a complex life cycle and produce remarkably resistant enzymes such as esterases (Essoussi *et al.*, 2010; Jaouani *et al.*, 2012;

Normand *et al.*, 2014). They also have the ability to resist adverse environmental conditions such as ultraviolet light, ionizing radiation, desiccation and heavy metals (Rainey *et al.*, 2005; Gtari *et al.*, 2012; Montero-Calasanz *et al.*, 2014, 2015). This resistance to environmental hazards represents a trait of Terrabacteria, a well-supported phylogenetic group composed of Actinobacteria and four other major lineages of eubacteria (Firmicutes, Cyanobacteria, Chloroflexi and *Deinococcus-Thermus*) that colonized land 3.05–2.78 Ga (Tunnacliffe and Lapinski, 2003; Battistuzzi *et al.*, 2004; Battistuzzi and Hedges, 2009).

Surprisingly, Geodermatophilaceae are present in a variety of biotopes including prominently rocks (Eppard *et al.*, 1996) and desert sandy soils (Montero-Calasanz *et al.*, 2012, 2013, 2013a, 2013c, 2013d, 2013e, 2013f; Liu *et al.*, 2014). Although considered endemic to soils, evolution of

Geodermatophilaceae has continued in specialized land biotopes. Indeed, soil and stone niches have yielded a wealth of knowledge regarding the extant distribution of Geodermatophilaceae (Gtari *et al.*, 2012; Normand *et al.*, 2014), raising questions about their evolution and the mechanisms of adaptation to harsh environments.

After their uplift by storms, Geodermatophilaceae have the potential to travel thousands of kilometers in the atmosphere (Chuvochina *et al.*, 2011). Consequently, stone surfaces can be colonized by these wind-borne microbes (Essoussi *et al.*, 2012). These surfaces are often covered with growth (called 'patinas', 'varnish-like' or 'tintenstriches'), which comprises complex communities of eukaryotes and prokaryotes, recurrent among which are Actinobacteria (Eppard *et al.*, 1996; Kuhlman *et al.*, 2006) including Geodermatophilaceae (Urzi *et al.*, 2001). With regard to biopitting, one hypothesis is that acid secretion and high carbon dioxide (CO_2) emissions from combustion engines result in alternating episodes of calcareous solubilization and precipitation. The microbial communities located in biopits have been analyzed by microbiological and molecular methods, and has been found to be complex, with a recurrence of Geodermatophilaceae. To deduce strategies for the dispersal of established biofilms and propose restoration approaches, identification of the components of the matrix of these biopolymers is required (Nijland *et al.*, 2010).

To understand how Geodermatophilaceae adapt to stones and soil, here we present a proteogenome analysis and detailed comparison of *B. saxobsidens* (Bs) (Chouaia *et al.*, 2012), *M. marinus* (Mm) (Normand *et al.*, 2012) and *G. obscurus* (Go) (Ivanova *et al.*, 2010). To get a close-up view of their physiology, we analyzed the proteome content of stationary-phase Bs, Mm and Go cells by a liquid chromatography–tandem mass spectrometry (MS/MS) shotgun approach and semi-quantification by spectral counting. We focussed on the identification of interesting proteic biomarkers of potential physiological value. Bs, Mm and Go are the first Geodermatophilaceae whose genomes and proteomes have been analyzed jointly. Now, they represent new models of choice for studies of niche adaptation among Terrabacteria.

Materials and methods

Bioinformatics approaches
Genes were classed into Clusters of Orthologous Groups (COG) (Tatusov *et al.*, 2001) and were retrieved from the Mage platform (Vallenet *et al.*, 2006). Metabolic pathways were analyzed using BioCyc (Caspi *et al.*, 2010). Identification of duplicated, lost or horizontally transferred genes was done using the Mage platform (phyloprofile) as was the identification of the core and extended genome. Clustered Regularly Interspaced Short

Palindromic Repeats were identified with the CRISPI database http://crispi.genouest.org/ (Rousseau *et al.*, 2009). Genomic islands were identified with Island-Viewer available at http://www.pathogenomics.sfu.ca/islandviewer/query.php (Langille and Brinkman, 2009). Phylogenetic analysis was done using MEGA6 (Tamura *et al.*, 2013) and the inferred topology was drawn and integrated with the genomic context at the Microbial Genomic context Viewer accessible at http://mgcv.cmbi.ru.nl/ (Overmars *et al.*, 2013).

Correspondence analysis was done as previously described (Benzécri, 1973) using the R software (R Development Core Team, 2007) on COG numbers (obtained from the Mage platform/Genomic Tools) and on numbers of coding DNA sequences (CDS) containing transcription and signaling domains (obtained from the Mage platform/Search Interpro) as previously described (Santos *et al.*, 2009).

Bacterial growth and proteome sample preparation
Cells of DD2 (Bs), BC501 (Mm) and DSM 43160 (Go) were plated onto Luedemann medium (yeast extract, malt extract, glucose, soluble starch and calcium carbonate ($CaCO_3$)) and incubated for 72 h at 28 °C as described previously (Luedemann, 1968). Bacterial cultures (15 mg wet weight) were resuspended into 90 µl of lithium dodecyl sulfate β-mercaptoethanol protein gel sample buffer (Invitrogen, Carlsbad, CA, USA) and incubated at 99 °C for 5 min as indicated previously (Hartmann *et al.*, 2014). Before SDS-polyacrylamide gel electrophoresis analysis on 10% Bis-Tris NuPAGE gels (Invitrogen), the samples were briefly centrifuged to remove large aggregates. A volume of 20 µl of the proteome of Bs, Mm and Go (corresponding to 160 µg of total proteins) was loaded per well. Three independent biological replicates were analyzed per microorganism. Sodium dodecyl sulfate-PAGE was carried out in 1 × 3-(*N*-morpholino) propanesulfonic acid solution (Invitrogen) on a XCell SureLock Mini-cell (Invitrogen) under a constant voltage of 200 V for 5 min. Gels were stained with SimplyBlue SafeStain, a ready-to-use Coomassie G-250 stain (Invitrogen). SeeBlue Plus2 (Invitrogen) was used as a molecular weight marker. Polyacrylamide gel bands (equivalent in volume to 50 µl) comprising the entire proteomes— one band per entire proteome—were cut and processed for in-gel proteolysis with Trypsin Sequencing Grade (Roche, Meylan, France) followed by the ProteaseMax protocol (Promega, Madison, WI, USA) as described previously (Clair *et al.*, 2010).

Nano liquid chromatography–MS/MS analysis
Peptide digests were resolved on an Ultimate 3000 LC systerm, Dionex-LC Packings (Thermo-Scientific, Villebon-sur-Yvette, France) before MS/MS measurements were done with a LTQ-Orbitrap XL (Thermo-Scientific) as described previously (Dedieu *et al.*, 2011). MS/MS spectra were processed and

interpreted with the MASCOT 2.3.02 search engine (Matrix Science, London, UK) with standard parameters as indicated previously (Hartmann and Armengaud, 2014) against databases corresponding to a complete list of annotated CDS from either Bs (NCBI RefSeq: NC_016943.1), Mm (BioProject: PRJEA167487, PRJEA82845) or Go (BioProject: PRJNA43725, PRJNA29547). Peptide matches with a score above their peptidic identity threshold were filtered at $P < 0.05$. A protein was only validated when at least two peptides had been assigned. Using a previously described approach (Liu et al., 2004; Zivanovic et al., 2009), protein abundance was evaluated by shotgun analysis using MS/MS spectral counts. Normalized spectral count abundance factors were calculated (Paoletti et al., 2006). The sum of all normalized spectral count abundance factors — 100%—was calculated for each bacterium: Bs (142.63), Mm (46.38) and Go (162.87). Accordingly, all values in this study were separated from the locus tags with a comma or between parentheses and represent the normalized spectral count abundance factor percentages, unless otherwise stipulated. The MS proteomics data have been deposited at the open access library of ProteomeXchange Consortium (http://www.proteomexchange.org/) (Vizcaino et al., 2014) with the data set identifiers PXD001519, PXD001518 and PXD001520 for Bs, Mm and Go, respectively.

Results and Discussion

Characteristics of proteogenomes

Life in biotopes with low trophic resources has driven the three Geodermatophilaceae members toward medium-sized genomes (Ivanova et al., 2010; Chouaia et al., 2012; Normand et al., 2012) from 4.87 to 5.32 Mb (Figure 1 and Supplementary Table S1). The three plasmidless genomes had very high G+C% (72.95–74.1%). Under unstressed conditions, the three proteomes were analyzed by a high-throughput shotgun procedure (Christie-Oleza and Armengaud, 2010). For Bs (PXD001519, 39889 MS/MS spectra, Supplementary Data S1), Mm (PXD001518, 14729 MS/MS spectra, Supplementary Data S2) and Go (PXD001520, 14829 MS/MS spectra, Supplementary Data S3), 5506, 1940 and 6884 spectra could be assigned to 553, 100 and 370 proteins, respectively. These three data sets represent the first proteogenome references for Geodermatophilaceae. Figure 1 depicts the proteins detected in this study.

We have predicted 3277 genes (30% amino acid identity) shared by the 3 genomes, which were sorted into 7 possible and mutually exclusive Venn groups (Supplementary Figure S1). Significant similarity to previously reported genes of known function allowed us to assign a putative function to 3231, 3643 and 3351 protein-coding genes in Bs, Mm and Go, respectively. The remaining genes were designated 'proteins with unknown function'. The analysed proteomes of Bs, Mm and Go indicate the expression of four, two and three such 'proteins of unknown function' (Supplementary Table S2), among the most highly expressed proteins—55, 9 and 42 as summarized in Supplementary Figure S2—that account for half of the total number of assigned spectra, respectively. As will become apparent below, some of these 'proteins of unknown function' seem to have a primordial role in niche adaptation of the host bacteria.

The three genomes have 70% of their CDS that could be ascribed to the COG category (Supplementary Table S3) (Tatusov et al., 2001; Vallenet et al., 2006). Correspondance analysis showed that the three Geodermatophilaceae genomes are close to one another. All three, in particular Bs and Mm, have a high proportion of [T] (signal) and [P] (inorganic ion transport and metabolism) categories, which is evocative of a lifestyle in a mineral-rich biotope (Supplementary Figure S3A). The overall distribution of the COG
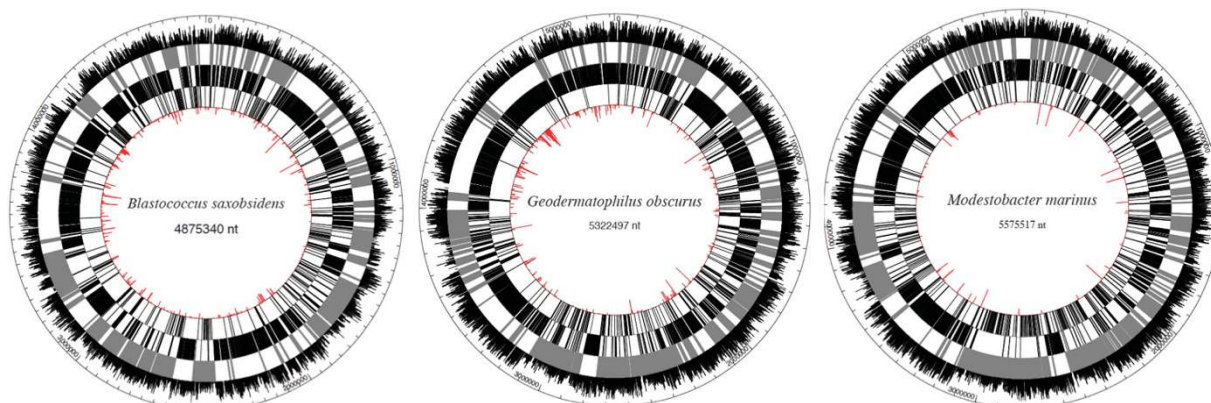

Figure 1 Circular representation of the three genomes with detected proteins. From the outside are 1, the coordinates; 2, the G+C% (ranging from 70 to 80%); 3, the horizontal gene transfer (HGT) predicted by the software RGP run on Mage (in grey); 4, the core genome or the genes shared by the three genomes (threshold of 50% identity over 80% of the length of the shorter sequence, present in a synton); 5, the genes specific to each genome (absent from the other two genomes, minus the 'genes of unknown function'; and 6, the proteins detected in this study (in red).

profile as well as the abundance of transcription factors and signaling molecules (Supplementary Figures S3B and C) constitute a signature that may be associated to lifestyle (Santos *et al.*, 2009). Compared with other Actinobacteria, the three genomes contained the highest absolute numbers of CDS containing PAS and EAL domains (Supplementary Figure S3C).

Our proteomic results showed that the most represented COG categories among the most highly expressed proteins are as follows: [J] translation, ribosomal structure and biogenesis (~33%), [C] energy production and conversion (~22%) and [I] lipid transport and metabolism (~9%) in Bs; [G] carbohydrate transport and metabolism (~33%), [E] amino acid transport and metabolism (~22%) and [P] (~22%) in Mm; and [J] (~33%), [C] (~26%) and [P] (~7%) in Go. The category [P] is represented by >5% of the most highly expressed proteins of Bs. These proteomic results are a further support for the presence of the monophyletic group composed of Bs and Go, which has been confirmed *in silico* by bioinformatics (Sen *et al.*, 2014) as well as *in vitro* by microbiological and biochemical markers (Normand *et al.*, 2014).

'Molecular tinkering/opportunism'strategies

Bs, Mm and Go genomes exhibit at least three strategies related to 'molecular tinkering/opportunism' (Jacob, 1977; Gogarten *et al.*, 2002; Laubichler, 2007) (Supplementary Table S4): (i) domain duplication; (ii) horizontal gene transfer, genes absent in two of the analyzed genomes but present in one of a group of more distant Actinobacteria; and (iii) rapid evolution, to create ORFans (Daubin and Ochman, 2004; Fukuchi and Nishikawa, 2004). Mm has 429 duplicated genes (7.9% of the genome), whereas in Bs and Go the number of these genes is slightly lower (representing 4.8% and 6.5%, respectively). Removal of a complex nutrient induced a motile state in these bacteria—motile budding rods called R-forms (Ishiguro and Wolfe, 1970,1974). The most highly expressed protein in Bs and Mm was flagellin synthesis, Lin (BLASA_0851, 4.96% and MODMU_1040, 11.51%) that has a paralog encoding a flagellar hook-associated protein FigL (BLASA_0855 and MODMU_1044). The same duplication event was observed in Go between similar paralogs, a flagellin (Gobs_0985, 0.02%) and a FigL (Gobs_0990). Rates of horizontal gene transfer amount to 6.9%, 8.9% and 6.8% in Bs, Mm and Go, respectively, which is consistant with hostile rock environments where antibiosis is often an unaffordable luxury (Friedmann and Ocampo-Friedmann, 1984). For example, our proteomic analyses indicated the presence of a transposase in Bs (BLASA_4384, 0.01%). The number of ORFans with 7.4–10.2% of the three genomes is much higher than the number found for *Escherichia coli* (3.5%) (Daubin and Ochman, 2004), a difference that could be linked to an unexpectedly higher rate of horizontal gene transfer in, on the surface and outside stones than in the

promiscuous gastrointestinal tract. Our proteomic approach allowed the identification of five, two and six ORFans that may have important functional roles in Bs, Mm and Go, respectively (Supplementary Table S2). A computation of lost genes—genes present in two genomes but absent in the third—shows that Bs, by far, had the highest number of lost genes (515 CDS) (Supplementary Table S4).

First-line defense strategies

Genome analyses indicate that Bs, Mm and Go possess several genes putatively involved in carotenoid biosynthesis (Supplementary Table S5). The Bs orange pigment absorbs at 230–270 and at 450–500 nm, whereas Mm and Go pigments are quite comparable and absorb almost continuously between 200 and 750 nm (Supplementary Figure S4). Bs has a putative operon (BLASA_0209–0214, *crtB2*, *hopC*, *ispA*, *shc*, *hpnH*, *ilvC*) that is absent in Mm and Go, and this could explain its intense orange pigment. Moreover, the expression of an uncharacterized enzyme involved in pigment biosynthesis (BLASA_3306, 0.05%) was detected in Bs.

The three genomes also possess impressive arrays of genes involved in stress relief, reactive oxygen species (ROS, superoxide anions ($O2^{\bullet-}$), hydrogen peroxide (H_2O_2) and hydroxyl radicals ($HO^{\bullet}$)) detoxification and DNA protection and repair (Supplementary Table S6). The Bs, Mm and Go proteomes express a nickel-containing superoxide dismutase (BLASA_3991, 0.83%; MODMU_4573, 0.57%; and Gobs_4176, 0.29%) and two catalases. Catalase (KatE) was one of the most highly expressed proteins in Mm (MODMU_2078, 4.64%). Its ortholog, Gobs_2125 (0.74%), belongs to the first 42 proteins that accounted for half of the total number of assigned spectra to the proteome of Go. Bs also highly expressed KatE (BLASA_3094, 0.54%). In addition, Bs expressed a manganese-containing catalase (KatA: BLASA_0196, 0.04%), but to a lesser extent than Go KatE (only 2 versus 75 spectral counts, when cumulating data from the triplicated samples). The transformation of the non-essential amino acid sarcosine, a source of carbon and energy derived from the osmoprotectant betaine, into the essential amino acid glycine generates H_2O_2. The sarcosine oxidase subunits (*soxB*, *soxD* and *soxA*) and carbon monoxide (CO) dehydrogenase subunit G (*coxG*) genes form an operon in ROS-resistant Mm and Go (MODMU_3072–MODMU_3075 and Gobs_2883–Gobs_2880, respectively), and were not detected in the ROS-sensitive Bs (Gtari *et al.*, 2012). Supplementary Table S7 lists other selected physiological features present in Bs, Mm and Go. For instance, in accordance with our previously published experimental data (Gtari *et al.*, 2012), orthologs of metal tolerance determinants (Janssen *et al.*, 2010) were detected in the three studied genomes.

Linear density of genomic DNA double-strand breaks inflicted per Gy per Mbp (0.002–0.004) is

similar for diverse bacteria (Daly (2009,2011) and references therein). Acute doses of 0.9, 6 and 9 kGy (Gtari et al., 2012) are predicted to inflict ~ 18, ~ 128 and ~ 192 double-strand breaks in Bs (~4.88 Mbp), Mm (~5.33 Mbp) and Go (~5.33 Mbp), respectively. Although absent in almost all actinobacterial species, all three genomes contained a non-homologous end joining operon, BLASA_3099–BLASA_3097 in Bs, MODMU_2074–MODMU_2076 in Mm and Gobs_2119–Gobs_2121 in Go, (Supplementary Figure S5). Bs possesses a supplementary putative Ku protein (BLASA_1744) that forms another operon with 'proteins of unknown function'. These findings suggest that non-homologous end joining may be a major pathway of double-strand breaks repair in Geodermatophilaceae.

Most abundant proteomic biomarkers and niche signatures

Bs was isolated from deep (2 cm), from stones found around the Mediterranean using a chisel and hammer, to eliminate the surface layers (Urzi et al., 2004), and it is predominant in the deeper fraction, that is, about 2 cm below the stone surface (Gtari et al., 2012). The proteomic analysis results (Supplementary Figure S2) suggest that unstressed Bs has evolved a survival strategy inside stones based on the following: (i) heavy investment in protein synthesis and in preventing their aggregation (ribosomal proteins and GroEL); (ii) detection and response to the changes of the environmental external nutrients (UDP glucose, phosphate and so on) concentrations (LysM, UshA and PhoU) (Buist et al., 2008; Marzan and Shimizu, 2011); (iii) scavenging ROS (superoxide dismutase); and (iv) transport of oxygen (hemerythrin). The presence of enzymes using anaerobic terminal electron acceptors in Bs, Mm and Go (Supplementary Table 7) indicates that formate (HCO_2^-) and nitrate (NO_3^-) anaerobic respiration may be possible.

We also discovered four highly expressed biomarkers (MODMU_0153, 5.83%; MODMU_0507, 5.41%; MODMU_1130, 3.92%; and MODMU_3547, 3.17%) of Mm, isolated from a white marble surface (Carrara, Tuscany, Italy) (Urzi et al., 2001) and predominant in the upper fraction (about 2 mm of stone surface) (Gtari et al., 2012; Supplementary Figure S2). Orthologs of these four highly expressed biomarkers are proteins implicated in the development of biofilms.

The outside of the stone has also been investigated (Urzi et al., 2001; Berdoulay and Salvado, 2009; Macedo et al., 2009). Growth on stone surfaces means either reliance on photosynthesis or on nutrients carried by the rain, air or through the stone itself. Concerning operons encoding genes for photosynthesis reactions, Go, isolated from soil of the Amargosa Desert (Nevada, USA) (Luedemann, 1968) contains three (Gobs_1696–Gobs_1703, Gobs_4550–Gobs_4544 and Gobs_4558–Gobs_4551) and Bs contains two (BLASA_0681 and BLASA_2555–

BLASA_2552). Surprisingly, Mm contains only a single NADPH-ferredoxin reductase (fprA) gene (MODMU_0890). Two genes encoding ribulose-1,5-bisphosphate carboxylase/oxygenasewere identified only in Go (Gobs_1448, 0.03% and Gobs_2026), suggesting that the strain obtains both carbon and energy via carboxydotrophy. In addition, Go is characterized by the presence of DNA-related biomarkers including a highly expressed DNA-binding histone-like protein (Gobs_0298, 1.13%) (Supplementary Figure S2). In contrast to Mm, both Go and Bs have many similar highly expressed proteomic biomarkers such as a Dps-like iron-chelating protein (Gobs_3661 and BLASA_1121) that may limit, through the confinement of free iron, the Fenton-derived production of $HO^{\bullet}$ (Williams et al., 2007; Confalonieri and Sommer, 2011). Yet, Bs has more highly expressed biomarkers associated with the production of ROS (Supplementary Figure 2)— cytochromes, flavoproteins and so on—representing a cellular benchmark for the proclivity of cells to resist to stress such as ionizing radiation (Ghosal et al. (2005) and references therein). Given this current state of affairs, it comes as no surprise that Bs is ROS and ionizing radiation sensitive (Gtari et al., 2012).

Rain is known to carry nitrogen compounds (Singh and Agrawal, 2008) such as nitric acid as well as traces of sulfur (Raybould et al., 1977). Besides glutamine synthetase and glutamate synthase for ammonium assimilation, the three genomes contain a conserved operon coding for an ammonium transporter (AmtB) and a nitrogen regulatory protein (GlnB) (Supplementary Table S7). Air carries numerous volatile organic compounds, prominent among which is CO (Austin et al., 2001). Study of the genomes revealed that Bs, Mm and Go have several copies of the coxLMS operon (Supplementary Table S7), which would help them oxidize CO. Such multiple copies (Wu et al., 2005) are always an indication of a strong selective pressure (Oda et al., 2005; Lee et al., 2009). Expression of the proteins— CO dehydrogenase subunits and acetyl-coenzyme A synthetase—(Supplementary Table S2) corresponding to some of the identified orthologous genes (Supplementary Table S7) of the Wood–Ljungdahl pathway has been detected under standard growth conditions of Bs and Go, suggesting metabolic utility of this pathway. Contrarily to CoxM and CoxL, acetyl-coenzyme A synthetase protein was not detected in Mm (Supplementary Table S2). Thus, the three strains inhabit exacting biotopes, which necessitate a rich array of transport systems, storage components, a motility machinery and energy-generating pathways (Figure 2). These biomarkers have shed new light on the microniche signature for each rock-dwelling terrabacteria.

Conclusion and perspectives

Here, the complete genome sequences of three Geodermatophilaceae members, Bs, Mm and Go,
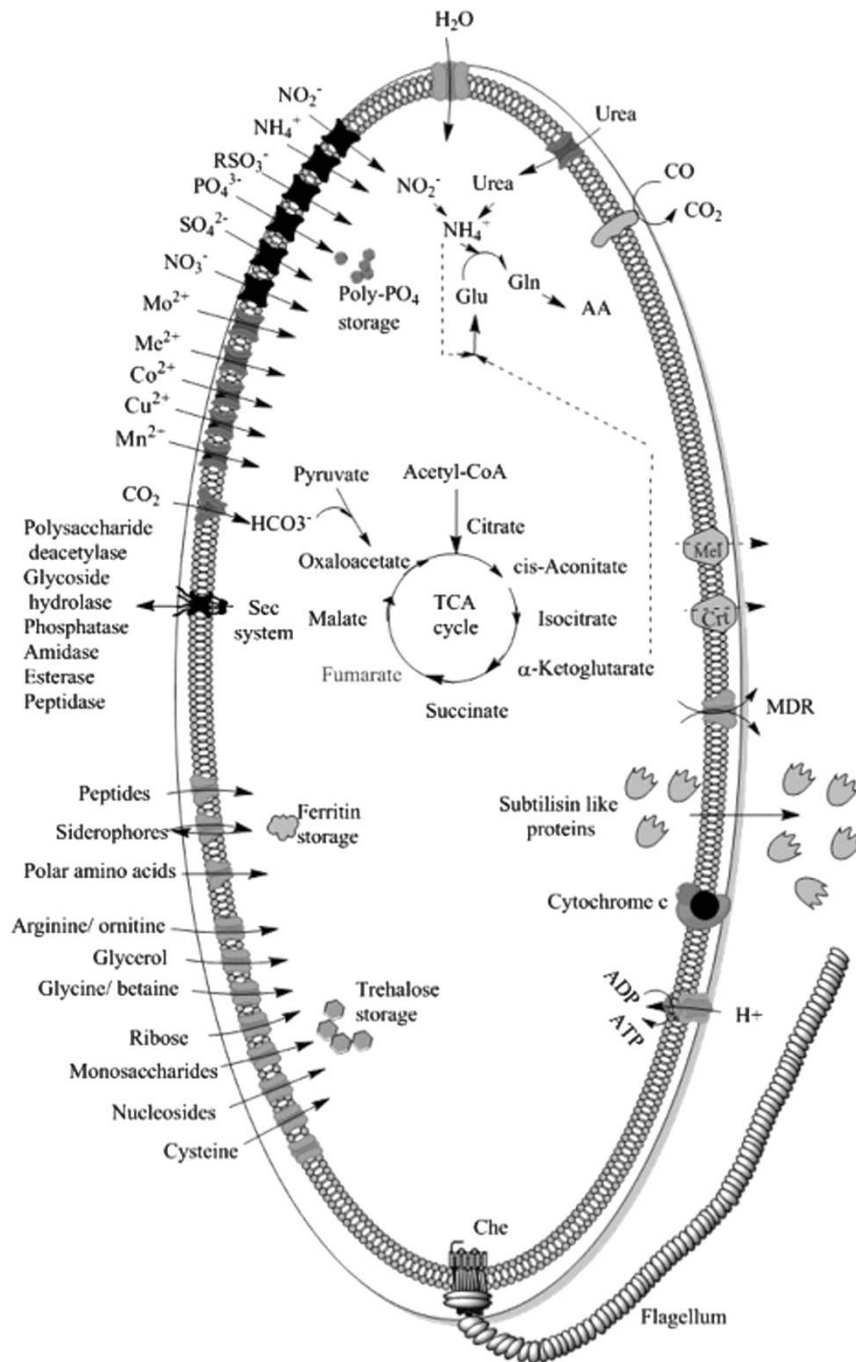
Figure 2 Schematic view of the Geodermatophilaceae physiological determinants. Transport systems, storage components, the motility machinery and the main energy generating pathways are represented.

with contrasted physiologies and ecological microniches (Normand *et al.*, 2014), together with the analysis of their proteomes under unstressed conditions, should help provide a solid foundation for investigating the varied strategies to adapt to their lifestyles. In particular, comparison of the three genomes provided an opportunity to analyze how Bs, Mm and Go can respond to stresses such as ROS mainly via pigmentation and catalase production and double-strand break through the

non-homologous end joining pathway. Moreover, highly expressed proteomic biomarkers of Bs, Mm and Go were depicted. The identification of thse biomarkers have shed new light on the physiological and biochemical traits that are unique to each species and its ecological microniche. In particular, the Mm exoproteome was almost as dominant as the cellular proteome, which hinders a deeper proteomic view (Armengaud *et al.*, 2012). Undoubtedly, much of the future progress in

studying Bs, Mm and Go rests squarely on the shoulders of research performed with their stressed proteogenomes.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

References

Ahrens R, Moll G. (1970). Ein neues knospendes Bakterium aus der Ostsee. [A new budding bacterium from the Baltic Sea]. *Arch Mikrobiol* **70**: 243–265.

Armengaud J, Christie-Oleza JA, Clair G, Malard V, Duport C. (2012). Exoproteomics: exploring the world around biological systems. *Expert Rev Proteomics* **9**: 561–575.

Austin CC, Wang D, Ecobichon DJ, Dussault G. (2001). Characterization of volatile organic compounds in smoke at experimental fires. *J Toxicol Environ Health A* **63**: 191–206.

Battistuzzi FU, Feijao A, Hedges SB. (2004). A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BMC Evol Biol* **4**: 44.

Battistuzzi FU, Hedges SB. (2009). A major clade of prokaryotes with ancient adaptations to life on land. *Mol Biol Evol* **26**: 335–343.

Benzécri J-P. (1973). *L'Analyse des Données* **Volume II**. L'Analyse des Correspondances: Paris, France Dunod-Paris, France Dunod.

Berdoulay M, Salvado JC. (2009). Genetic characterization of microbial communities living at the surface of building stones. *Lett Appl Microbiol* **49**: 311–316.

Buist G, Steen A, Kok J, Kuipers OP. (2008). LysM, a widely distributed protein motif for binding to (peptido) glycans. *Mol Microbiol* **68**: 838–847.

Caspi R, Altman T, Dale JM, Dreher K, Fulcher CA, Gilham F *et al.* (2010). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* **38**: D473–D479.

Chouaia B, Crotti E, Brusetti L, Daffonchio D, Essoussi I, Nouioui I *et al.* (2012). Genome sequence of *Blastococcus saxobsidens* DD2, a stone-inhabiting bacterium. *J Bacteriol* **194**: 2752–2753.

Christie-Oleza JA, Armengaud J. (2010). In-depth analysis of exoproteomes from marine bacteria by shotgun liquid chromatography-tandem mass spectrometry: the *Ruegeria pomeroyi* DSS-3 case-study. *Mar Drugs* **8**: 2223–2239.

Chuvochina MS, Marie D, Chevaillier S, Petit JR, Normand P, Alekhina IA *et al.* (2011). Community variability of bacteria in alpine snow (Mont Blanc) containing Saharan dust deposition and their snow colonisation potential. *Microbes Environ* **26**: 237–247.

Clair G, Roussi S, Armengaud J, Duport C. (2010). Expanding the known repertoire of virulence factors produced by *Bacillus cereus* through early secretome profiling in three redox conditions. *Mol Cell Proteomics* **9**: 1486–1498.

Confalonieri F, Sommer S. (2011). Bacterial and archaeal resistance to ionizing radiation. *J Phys Conf Ser* **26**10.1088/1742-6596/1261/1081/012005.

Daly MJ. (2009). A new perspective on radiation resistance based on *Deinococcus radiodurans*. *Nat Rev Microbiol* **7**: 237–245.

Daly MJ. (2011). Death by protein damage in irradiated cells. *DNA Repair (Amst)* **11**: 12–21.

Daubin V, Ochman H. (2004). Bacterial genomes as new gene homes: the genealogy of ORFans in *E. coli*. *Genome Res* **14**: 1036–1042.

Dedieu A, Gaillard JC, Pourcher T, Darrouzet E, Armengaud J. (2011). Revisiting iodination sites in thyroglobulin with an organ-oriented shotgun strategy. *J Biol Chem* **286**: 259–269.

Eppard M, Krumbein WE, Koch C, Rhiel E, Staley JT, Stackebrandt E. (1996). Morphological, physiological, and molecular characterization of actinomycetes isolated from dry soil, rocks, and monument surfaces. *Arch Microbiol* **166**: 12–22.

Essoussi I, Boujmil R, Nouioui I, Abbassi-Ghozzi I, Hamza A, Boudabous A *et al.* (2012). Genetic diversity and esterase-profiling of Actinobacteria isolated from Sahara desert stones and monuments. *Geomicrobiol J* **29**: 23–28.

Essoussi I, Ghodhbane-Gtari F, Amairi H, Sghaier H, Jaouani A, Brusetti L *et al.* (2010). Esterase as an enzymatic signature of *Geodermatophilaceae* adaptability to Sahara desert stones and monuments. *J Appl Microbiol* **108**: 1723–1732.

Friedmann EI, Ocampo-Friedmann RO. (1984). Endolithic microorganisms in extreme dry environments: analysis of a lithobiontic habitatIn:Klug MJ, Reddy CA (eds) *Current Perspectives in Microbial Ecology. Proc. Third Internat. Symp. Microb. Ecol., Am. Soc. Microbiol.*, Washington, DCpp 177–185.

Fukuchi S, Nishikawa K. (2004). Estimation of the number of authentic orphan genes in bacterial genomes. *DNA Res* **11**: 219–231.

Ghosal D, Omelchenko MV, Gaidamakova EK, Matrosova VY, Vasilenko A, Venkateswaran A *et al.* (2005). How radiation kills cells: survival of *Deinococcus radiodurans* and *Shewanella oneidensis* under oxidative stress. *FEMS Microbiol Rev* **29**: 361–375.

Gogarten JP, Doolittle WF, Lawrence JG. (2002). Prokaryotic evolution in light of gene transfer. *Mol Biol Evol* **19**: 2226–2238.

Gtari M, Essoussi I, Maaoui R, Sghaier H, Boujmil R, Gury J *et al.* (2012). Contrasted resistance of stone-dwelling *Geodermatophilaceae* species to stresses known to give rise to reactive oxygen species. *FEMS Microbiol Ecol* **80**: 566–577.

Hartmann EM, Allain F, Gaillard JC, Pible O, Armengaud J. (2014). Taking the shortcut for high-throughput shotgun proteomic analysis of bacteria. *Methods Mol Biol* **1197**: 275–285.

Hartmann EM, Armengaud J. (2014). Shotgun proteomics suggests involvement of additional enzymes in dioxin degradation by *Sphingomonas wittichii* RW1. *Environ Microbiol* **16**: 162–176.

Ishiguro EE, Wolfe RS. (1970). Control of morphogenesis in *Geodermatophilus*: ultrastructural studies. *J Bacteriol* **104**: 566–580.

Ishiguro EE, Wolfe RS. (1974). Induction of morphogenesis in *Geodermatophilus* by inorganic cations and by organic nitrogenous cations. *J Bacteriol* **117**: 189–195.

Ivanova N, Sikorski J, Jando M, Munk C, Lapidus A, Glavina Del Rio T *et al.* (2010). Complete genome sequence of *Geodermatophilus obscurus* type strain (G-20 T). *Stand Genomic Sci* **2**: 158–167.

Jacob F. (1977). Evolution and tinkering. *Science* **196**: 1161–1166.

Janssen PJ, Van Houdt R, Moors H, Monsieurs P, Morin N, Michaux A *et al.* (2010). The complete genome sequence of *Cupriavidus metallidurans* strain CH34, a master survivalist in harsh and anthropogenic environments. *PLoS One* **5**: e10433.

Jaouani A, Neifar M, Hamza A, Chaabouni S, Martinez MJ, Gtari M. (2012). Purification and characterization of a highly thermostable esterase from the actinobacterium *Geodermatophilus obscurus* strain G20. *J Basic Microbiol* **52**: 653–660.

Kuhlman KR, Fusco WG, La Duc MT, Allenbach LB, Ball CL, Kuhlman GM *et al.* (2006). Diversity of microorganisms within rock varnish in the Whipple Mountains, California. *Appl Environ Microbiol* **72**: 1708–1715.

Langille MG, Brinkman FS. (2009). IslandViewer: an integrated interface for computational identification and visualization of genomic islands. *Bioinformatics* **25**: 664–665.

Laubichler MD. (2007). Tinkering: a conceptual and historical evaluation. *Novartis Found Symp* **284**: 20–29 discussion 29–34, 110–115.

Lee ZM, Bussema C 3rd, Schmidt TM. (2009). rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. *Nucleic Acids Res* **37**: D489–D493.

Liu H, Sadygov RG, Yates JR 3rd. (2004). A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* **76**: 4193–4201.

Liu R, Li K, Zhang H, Zhu J, Joshi D. (2014). Spatial distribution of microbial communities associated with dune landform in the Gurbantunggut Desert, China. *J Microbiol* **52**: 898–907.

Luedemann GM. (1968). *Geodermatophilus*, a new genus of the Dermatophilaceae (Actinomycetales). *J Bacteriol* **96**: 1848–1858.

Macedo MF, Miller AZ, Dionisio A, Saiz-Jimenez C. (2009). Biodiversity of cyanobacteria and green algae on monuments in the Mediterranean Basin: an overview. *Microbiology* **155**: 3476–3490.

Marzan LW, Shimizu K. (2011). Metabolic regulation of *Escherichia coli* and its *phoB* and *phoR* genes knock-out mutants under phosphate and nitrogen limitations as well as at acidic condition. *Microb Cell Fact* **10**: 39.

Mevs U, Stackebrandt E, Schumann P, Gallikowski CA, Hirsch P. (2000). *Modestobacter multiseptatus* gen. nov., sp. nov., a budding actinomycete from soils of the Asgard Range (Transantarctic Mountains). *Int J Syst Evol Microbiol* **50**: 337–346.

Montero-Calasanz MC, Goker M, Broughton WJ, Cattaneo A, Favet J, Potter G *et al.* (2013). *Geodermatophilus tzadiensis* sp. nov., a UV radiation-resistant bacterium isolated from sand of the Saharan desert. *Syst Appl Microbiol* **36**: 177–182.

Montero-Calasanz MC, Goker M, Potter G, Rohde M, Sproer C, Schumann P *et al.* (2013a). *Geodermatophilus*

africanus sp. nov., a halotolerant actinomycete isolated from Saharan desert sand. *Antonie Van Leeuwenhoek* **104**: 207–216.

Montero-Calasanz MC, Goker M, Potter G, Rohde M, Sproer C, Schumann P *et al.* (2012). *Geodermatophilus arenarius* sp. nov., a xerophilic actinomycete isolated from Saharan desert sand in Chad. *Extremophiles* **16**: 903–909.

Montero-Calasanz MC, Goker M, Potter G, Rohde M, Sproer C, Schumann P *et al.* (2013c). *Geodermatophilus normandii* sp. nov., isolated from Saharan desert sand. *Int J Syst Evol Microbiol* **63**: 3437–3443.

Montero-Calasanz MC, Goker M, Potter G, Rohde M, Sproer C, Schumann P *et al.* (2013d). *Geodermatophilus saharensis* sp. nov., isolated from sand of the Saharan desert in Chad. *Arch Microbiol* **195**: 153–159.

Montero-Calasanz MC, Goker M, Potter G, Rohde M, Sproer C, Schumann P *et al.* (2013e). *Geodermatophilus telluris* sp. nov., an actinomycete isolated from Saharan desert sand. *Int J Syst Evol Microbiol* **63**: 2254–2259.

Montero-Calasanz MC, Goker M, Rohde M, Schumann P, Potter G, Sproer C *et al.* (2013f). *Geodermatophilus siccatus* sp. nov., isolated from arid sand of the Saharan desert in Chad. *Antonie Van Leeuwenhoek* **103**: 449–456.

Montero-Calasanz MC, Hofner B, Goker M, Rohde M, Sproer C, Hezbri K *et al.* (2014). *Geodermatophilus poikilotrophi* sp. nov.: a multitolerant actinomycete isolated from dolomitic marble. *Biomed Res Int* **2014**: 11.

Montero-Calasanz MD, Hezbri K, Goker M, Sghaier H, Rohde M, Sproer C *et al.* (2015). Description of gamma radiation-resistant *Geodermatophilus dictyosporus* sp. nov. to accommodate the not validly named *Geodermatophilus obscurus* subsp. *dictyosporus* (Luedemann, 1968). *Extremophiles* **19**: 77–85.

Nijland R, Hall MJ, Burgess JG. (2010). Dispersal of biofilms by secreted, matrix degrading, bacterial DNase. *PLoS One* **5**: e15668.

Normand P. (2006). *Geodermatophilaceae* fam. nov., a formal description. *Int J Syst Evol Microbiol* **56**: 2277–2278.

Normand P, Daffonchio D, Gtari M. (2014)Family GeodermatophilaceaeIn:Rosenberg E(ed.) *The prokaryotes – Actinobacteria.* Springer-Verlag: Berlin, Heidelberg-Berlin, Heidelberg.

Normand P, Gury J, Pujic P, Chouaia B, Crotti E, Brusetti L *et al.* (2012). Genome sequence of radiation-resistant *Modestobacter marinus* strain BC501, a representative actinobacterium that thrives on calcareous stone surfaces. *J Bacteriol* **194**: 4773–4774.

Oda Y, Samanta SK, Rey FE, Wu L, Liu X, Yan T *et al.* (2005). Functional genomic analysis of three nitrogenase isozymes in the photosynthetic bacterium *Rhodopseudomonas palustris*. *J Bacteriol* **187**: 7784–7794.

Overmars L, Kerkhoven R, Siezen RJ, Francke C. (2013). MGcV: the microbial genomic context viewer for comparative genome analysis. *BMC Genomics* **14**: 209.

Paoletti AC, Parmely TJ, Tomomori-Sato C, Sato S, Zhu D, Conaway RC *et al.* (2006). Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. *Proc Natl Acad Sci USA* **103**: 18928–18933.

R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. R Foundation

for Statistical Computing, Vienna: AustriaAustriahttp://www.R-project.org.

Rainey FA, Ray K, Ferreira M, Gatz BZ, Nobre MF, Bagaley D *et al.* (2005). Extensive diversity of ionizing-radiation-resistant bacteria recovered from Sonoran Desert soil and description of nine new species of the genus *Deinococcus* obtained from a single soil sample. *Appl Environ Microbiol* **71**: 5225–5235.

Raybould CC, Unsworth MH, Gregory PJ. (1977). Sources of sulphur in rain collected below a wheat canopy. *Nature* **267**: 146–147.

Rousseau C, Gonnet M, Le Romancer M, Nicolas J. (2009). CRISPI: a CRISPR interactive database. *Bioinformatics* **25**: 3317–3318.

Santos CL, Tavares F, Thioulouse J, Normand P. (2009). A phylogenomic analysis of bacterial helix-turn-helix transcription factors. *FEMS Microbiol Rev* **33**: 411–429.

Sen A, Daubin V, Abrouk D, Gifford I, Berry AM, Normand P. (2014). The phylogeny of actinobacteria revisited in the light of complete genomes, the orders Frankiales and Micrococcales should be split into coherent entities. Proposal of *Frankiales* ord. nov., *Geodermatophilales* ord. nov., *Acidothermales* ord. nov. and *Nakamurellales* ord. nov. *Int J Syst Evol Microbiol* **64**: 3821–3832.

Singh A, Agrawal M. (2008). Acid rain and its ecological consequences. *J Environ Biol* **29**: 15–24.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**: 2725–2729.

Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS *et al.* (2001). The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res* **29**: 22–28.

Tunnacliffe A, Lapinski J. (2003). Resurrecting Van Leeuwenhoek's rotifers: a reappraisal of the role of disaccharides in anhydrobiosis. *Philos Trans R Soc Lond B Biol Sci* **358**: 1755–1771.

Urzi C, Brusetti L, Salamone P, Sorlini C, Stackebrandt E, Daffonchio D. (2001). Biodiversity of *Geodermatophilaceae* isolated from altered stones and monuments in the Mediterranean basin. *Environ Microbiol* **3**: 471–479.

Urzi C, Salamone P, Schumann P, Rohde M, Stackebrandt E. (2004). *Blastococcus saxobsidens* sp. nov., and emended descriptions of the genus *Blastococcus* Ahrens and Moll 1970 and *Blastococcus aggregatus* Ahrens and Moll 1970. *Int J Syst Evol Microbiol* **54**: 253–259.

Vallenet D, Labarre L, Rouy Z, Barbe V, Bocs S, Cruveiller S *et al.* (2006). MaGe: a microbial genome annotation system supported by synteny results. *Nucleic Acids Res* **34**: 53–65.

Vizcaino JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Rios D *et al.* (2014). ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat Biotechnol* **32**: 223–226.

Williams E, Lowe TM, Savas J, DiRuggiero J. (2007). Microarray analysis of the hyperthermophilic archaeon *Pyrococcus furiosus* exposed to gamma irradiation. *Extremophiles* **11**: 19–29.

Wu M, Ren Q, Durkin AS, Daugherty SC, Brinkac LM, Dodson RJ *et al.* (2005). Life in hot carbon monoxide: the complete genome sequence of *Carboxydothermus hydrogenoformans* Z-2901. *PLoS Genet* **1**: e65.

Zhou J, Rudd KE. (2013). EcoGene 3.0. *Nucleic Acids Res* **41**: D613–D624.

Zivanovic Y, Armengaud J, Lagorce A, Leplat C, Guerin P, Dutertre M *et al.* (2009). Genome analysis and genome-wide proteomics of *Thermococcus gammatolerans*, the most radioresistant organism known amongst the Archaea. *Genome Biol* **10**: R70.

Supplementary Information accompanies this paper on The ISME Journal website (http://www.nature.com/ismej)