

Genomics update

Some (bacilli) like it hot: genomics of *Geobacillus* species**David J. Studholme****Biosciences, University of Exeter, Geoffrey Pope Building, Stocker Road, Exeter EX4 4QD, UK.***What are *Geobacillus*?**

The genus *Geobacillus* includes thermophilic Gram-positive spore-forming bacteria that form a phylogenetically coherent clade within the family *Bacillaceae*. They are of great interest for biotechnology (as discussed below). These thermophiles seem to be ubiquitous; viable *Geobacillus* spores can be isolated in large quantities not only from hot environments such as hydrothermal vents, but also, paradoxically, from cool soils and cold ocean sediments (Zeigler, 2014).

These bacteria were previously categorized as 'Group 5' within the genus *Bacillus* but were subsequently split into the new genus *Geobacillus* (Nazina *et al.*, 2001). Many *Geobacillus* strains were previously described as belonging to a single species *Bacillus stearothermophilus*, but it was clear that there was great heterogeneity in physiology, preferred temperature range and other phenotypic characteristics among these strains. For example, see Fig. 1 showing three distinct colony morphologies among three strains described as '*B. stearothermophilus*'. It is now absolutely clear that there are several distinct species within *Geobacillus* and these can be distinguished by both genotype and phenotype (Nazina *et al.*, 2001; Banat *et al.*, 2004; Zeigler, 2005; Dinsdale *et al.*, 2011; Coorevits *et al.*, 2012).

Why are *Geobacillus* species of interest for biotechnology?

Geobacillus spp. are of interest for biotechnology as source of thermostable enzymes and natural products, digesters of lignocellulose, bioremediators of hydrocar-

bons, producers of bio-fuel, cellular factories for heterologous expression of enzymes and as hosts for directed evolution (Wiegel *et al.*, 1985; Niehaus *et al.*, 1999; Couñago and Shamoo, 2005; Marchant *et al.*, 2006; Cripps *et al.*, 2009; Taylor *et al.*, 2009; Tabachnikov and Shoham, 2013). Industrially important enzymes originating from *Geobacillus* spp. include lipases (Schmidt-Dannert *et al.*, 1998), glycoside hydrolases (Fridjonsson *et al.*, 1999; Bartosiak-Jentys *et al.*, 2013; Suzuki *et al.*, 2013), N-acylhomoserine lactonase (Seo *et al.*, 2011) and DNA polymerase I (Sandalli *et al.*, 2009) and protease (Chen *et al.*, 2004) among others. The advantages of using thermophilic bacteria as whole-cell biocatalysts were recently discussed in this journal (Taylor *et al.*, 2011) and include reduced risk of contamination, acceleration of biochemical processes and easier maintenance of anaerobic conditions. These bacteria also tend to ferment a wide range of substrates, utilizing both cellobiose and pentose sugars. In the context of bioethanol production, there is the additional advantage of reduced cooling costs and easier removal and recovery of the volatile product by sparging or partial vacuum thus also avoiding ethanol poisoning of the bacteria (Taylor *et al.*, 2009). Less positively, *Geobacillus* spp. are common contaminants in the dairy and food industries (Burgess *et al.*, 2010).

Which genomes have been sequenced?

At the time of writing (28 July 2014), 29 *Geobacillus* genome sequences are available (Table 1). These include representatives of all the major phylogenetic groups within the genus and include representatives of the species *G. thermoleovorans*, *G. kaustophilus*, *G. thermocatenu-latus*, *G. thermodenitrificans*, *G. stearothermophilus*, *G. caloxylosilyticus* and *G. thermoglucosidans* (formerly *G. thermoglucosidasius*) as well as several strains that have not been assigned to named species (Fig. 2). Genome sequences are also available for some other thermophilic members of the *Bacillaceae*, such as *Paenibacillus lautus* (Mead *et al.*, 2012) and *Bacillus coagulans* (Xu *et al.*, 2013) and for *Geobacillus*-infecting bacteriophage (Marks and Hamilton, 2014), but these will

*For correspondence. E-mail d.j.studholme@exeter.ac.uk; Tel. +44 (0) 1392 72 4678; Fax +44 (0) 1392 263434.
doi:10.1111/1751-7915.12161

Funding Information Work in my laboratory is currently supported by Biotechnology and Biological Sciences Research Council (BBSRC) grants BB/H016120/1, BB/1024631/1, BB/1025956/1, BB/K003240/2 and BB/L012499/1.

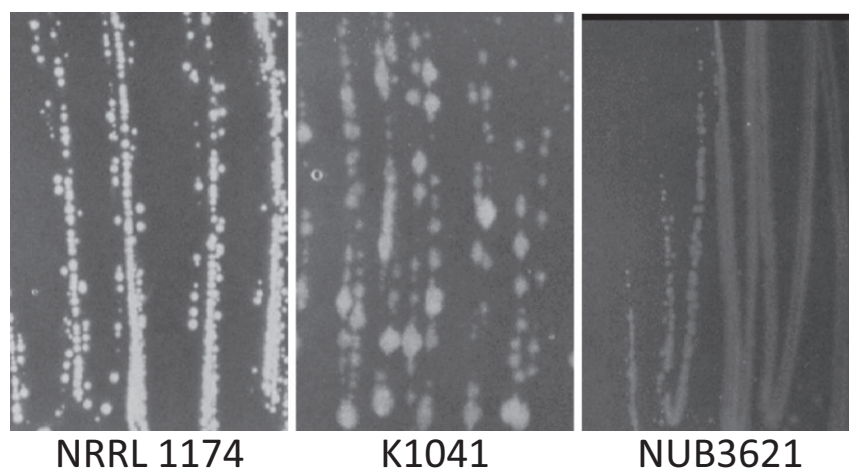


Fig. 1. Diverse colony morphologies of strains classified as '*G. stearothermophilus*'. Strains NRRL 1174, K1041 and NUB3621 were streaked-out on tryptic soy broth plates and incubated overnight at 50°C. Plates were photographed under identical conditions.

Table 1. *Geobacillus* strains whose genomes have been sequenced as of 26 July 2014.

Species and strain	Motivation for sequencing	Accession number	References
<i>G. caldxylosilyticus</i> CIC9	Not known	NZ_AMRO01000000.1	n. a.
<i>G. caldxylosilyticus</i> NBRC 107762	Not known	BAWO01000000.1	n. a.
<i>G. kaustophilus</i> GBlys	Lysogenic, containing an integrated prophage	NZ_BASG01000001.1	(Doi <i>et al.</i> , 2013)
<i>G. kaustophilus</i> HTA426	Source of novel glycoside hydrolases (6-phospho- β -glycosidase and β -fucosidase)	NC_006510.1	(Takami <i>et al.</i> , 2004)
<i>G. sp.</i> A8	Not known	NZ_AUXP01000001.1	n. a.
<i>G. sp.</i> C56-T3	Not known	NC_014206.1	n. a.
<i>G. sp.</i> CAMR12739	Hemicellulose degradation	JHUR01000001.1	(De Maayer <i>et al.</i> , 2014)
<i>G. sp.</i> CAMR5420	Hemicellulose degradation	JHUS01000001.1	(De Maayer <i>et al.</i> , 2014)
<i>G. sp.</i> FW23	Potential for degradation and utilization of oil (bioremediation of oil spills)	JGCJ01000001.1	(Pore <i>et al.</i> , 2014)
<i>G. sp.</i> G11MC16	Not known	NZ_ABVH01000001.1	n. a.
<i>G. sp.</i> GHH01	Source if thermostable and thermo-active secreted lipase	NC_020210.1	(Wiegand <i>et al.</i> , 2013)
<i>G. sp.</i> JF8	Degrades biphenyl and polychlorinated biphenyls (PCB)	NC_022080.4	(Shintani <i>et al.</i> , 2014)
<i>G. sp.</i> MAS1	Potential source of useful enzyme-encoding genes	NZ_AYSF01000001.1	(Siddiqui <i>et al.</i> , 2014)
<i>G. sp.</i> WCH70	Not known	NC_012793.1	n. a.
<i>G. sp.</i> WSUCF1	Abel to grow on lignocellulosic substrates	NZ_ATCO01000001.1	(Bhalla <i>et al.</i> , 2013)
<i>G. sp.</i> Y4.1MC1	Not known	NC_014650.1	n. a.
<i>G. sp.</i> Y412MC52	Not known	NC_014915.1	n. a.
<i>G. sp.</i> Y412MC61	Not known	NC_013411.1	n. a.
<i>G. stearothermophilus</i> ATCC 7953	Not known	JALS01000001.1	n. a.
<i>G. stearothermophilus</i> NUB3621	Genetically amenable host strain for metabolic engineering	AOTZ01000001.1	(Blanchard <i>et al.</i> , 2014)
<i>G. thermocatenulatus</i> GS-1	Not known	JFHZ01000001.1	n. a.
<i>G. thermodenitrificans</i> NG80-2	Denitrification and degradation of long-chain alkanes, facilitating oil recovery in oil reservoirs	NC_009328.1	(Feng <i>et al.</i> , 2007)
<i>G. thermodenitrificans</i> subsp. <i>thermodenitrificans</i> DSM 465	Comparative genomics between the alkane-utilizing NG80-2 and this strain which is unable to utilize alkanes	NZ_AYKT01000001.1	(Yao <i>et al.</i> , 2013)
<i>G. thermoglucosidans</i> TNO-09.020	Contaminant in dairy-processing environment	NZ_CM001483.1	(Zhao <i>et al.</i> , 2012)
<i>G. thermoglucosidasius</i> C56-YS93	Not known	NC_015660.1	n. a.
<i>G. thermoglucosidasius</i> NBRC 107763	Not known	BAWP01000001.1	n. a.
<i>G. thermoleovorans</i> B23 DNA	Alkane degrader with unidentified alkane monooxygenase	BATY01000001.1	(Boonmak <i>et al.</i> , 2013)
<i>G. thermoleovorans</i> CCB_US3_UF5	Not known	NC_016593.1	(Muhd Sakaff <i>et al.</i> , 2012)

Names are given as found in the GenBank sequence database. n.a., not available.

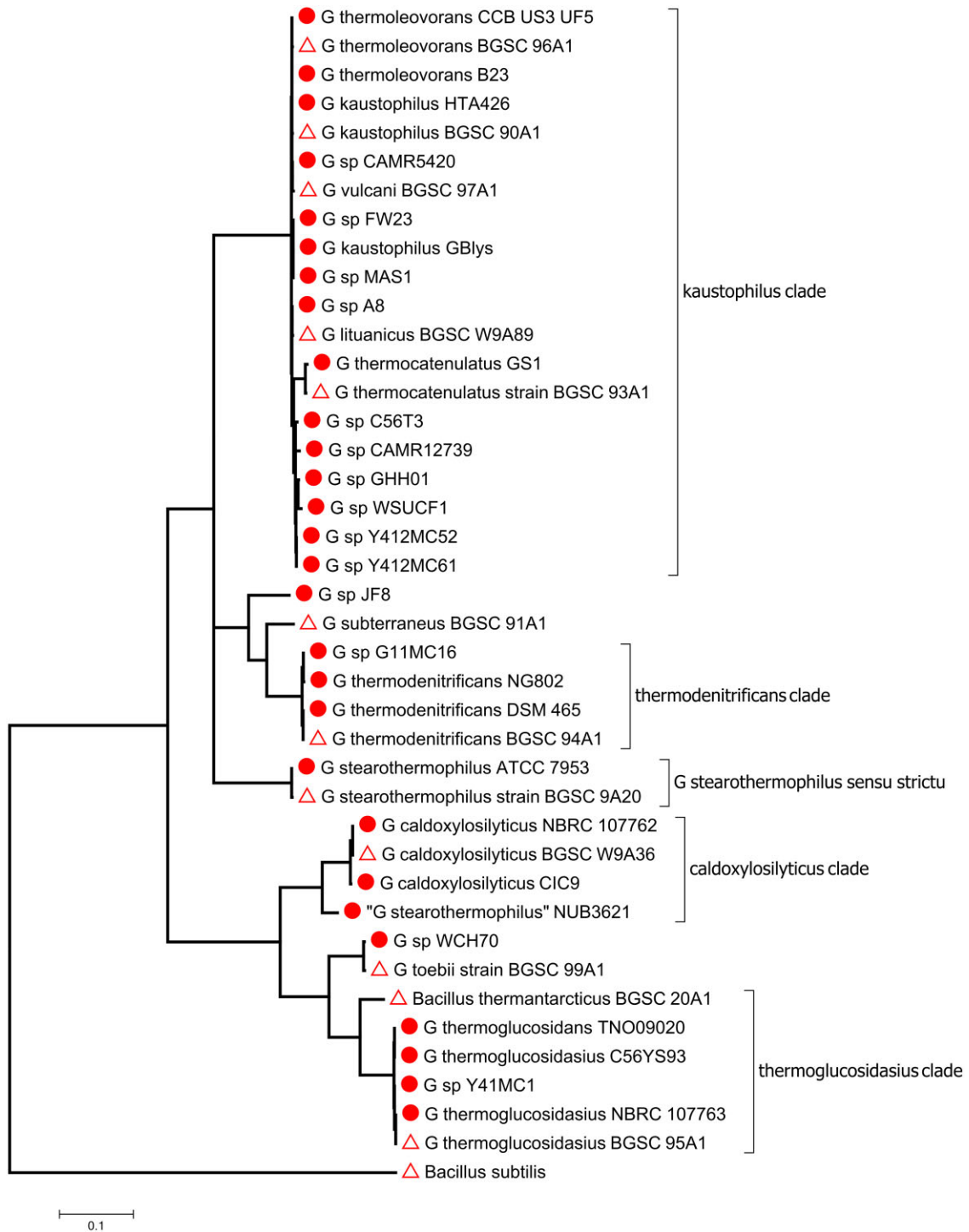


Fig. 2. Phylogenetic relationships among sequenced strains of *Geobacillus* inferred from a multiple sequence alignment of *recN* sequences. The circles indicate strains whose genomes have been sequenced, as listed in Table 1. The triangles indicate type strains of the various *Geobacillus* species; *recN* sequences from these are taken from a previous phylogenetic analysis by Zeigler (2005). The maximum-likelihood tree was generated using MEGA6 (Tamura *et al.*, 2013).

not be discussed here. The team who sequenced the genome of *Geobacillus* sp. MAS1 described this strain as '*G. thermopakistaniensis*', but this is not a validly named species and no justification was provided for its proposal as

a new species (Siddiqui *et al.*, 2014). On the basis of its *recN* sequence, a useful phylogenetic marker for *Geobacillus* spp. (Zeigler, 2005), strain MAS1 is closely related to the type strains of *G. kaustophilus* and

4 Genomics update

G. thermoleovorans (Fig. 2). Strain NUB3621 was described as '*G. stearothermophilus*' but as has been previously noted (Studholme *et al.*, 1999; Zeigler, 2005; Blanchard *et al.*, 2014), this strain is phylogenetically distinct from *B. stearothermophilus sensu strictu* and is more closely related to *G. caldoxylsilyticus* and, to a lesser extent, *G. thermoglucosidans* (Fig. 2). For more than half of the sequenced genomes, papers have been published describing and/or announcing the sequence data and usually indicating the particular features of the strain that motivated its sequencing. An insightful discussion of the biological lessons from *Geobacillus* genomes was previously published earlier this year, including surveys of genes involved in breakdown of plant-derived lignocellulose (Zeigler, 2014); but at that time, only 10 genome sequences were available.

The phylogenetic group within *Geobacillus* most richly represented by genome sequences is the clade containing *G. thermoleovorans*, *G. kaustophilus* and *G. thermocatenulatus* (see the 'kaustophilus clade' in Fig. 2). Based solely of sequences of the *recN* phylogenetic marker, it is not possible to precisely resolve relationships among sequenced strains within this group (Fig. 2). However, the availability of complete genome sequence data enables phylogenetic analysis based on single-nucleotide variants over the entire core genome, offering much greater resolution (Fig. 3A). According to the core-genome-wide phylogenetic analysis, the two strains assigned as *G. kaustophilus* do not form a phylogenetically coherent monophyletic clade. On the other hand, the two strains of *G. thermoleovorans* are closely related and share 99.4% nucleotide sequence identity [based on MUMMER2 alignments (Delcher *et al.*, 2002)]. Strain FW23 also appears to fall within this clade and, subject to phenotypic characterization, can probably be considered a member of this species too. *Geobacillus thermocatenulatus* GS-1 is much more divergent, sharing only 94% to 95% identity with the other strains in the clade, which is consistent with the *recN*-based analysis (Fig. 2). Strains Y412MC52 and YP412MC61 appear to be extremely closely related to each other, sharing 99.8% sequence identity and showing no detectable differences in gene content. Nucleotide sequence identities between clades are much lower; between *G. kaustophilus* and *G. thermoglucosidans*, there is approximately 84% identity.

The considerable amount of reticulation in the phylogenetic network (Fig. 3A) suggests significant horizontal genetic transfer within and among these species. This is further illustrated by the extent of variation in the variable component of the genome (Fig. 3B). Out of 3887 genes on the chromosome of *G. thermoleovorans* CCB US3 UF5, a total of 931 (approximately 24%) are variable (that is, they are absent from at least one of the other

sequenced genomes). The global pattern of gene content (Fig. 3B) broadly reflects the phylogenetic relationships (Fig. 3A): according to gene content, the genomes fall into four main clusters, indicated by four different colours of shading in Fig. 3B, which correspond to four zones of the phylogenetic network, shaded with the same colours in Fig. 3A. However, there are numerous genes whose distribution across the genomes is incongruent with core-genome phylogeny, again suggesting extensive horizontal transfer.

What benefits has the sequencing of *Geobacillus* genomes brought?

The availability of complete *Geobacillus* genome sequences has enabled or accelerated the discovery, cloning and exploitation of natural products. For example, the availability of the NG80-2 genome sequence (Feng *et al.*, 2007) enabled the discovery of thermostable homologues of the lantibiotic nisin in *G. thermodenitrificans* (Begley *et al.*, 2009; Garg *et al.*, 2012), opening the possibility of replacing nisin as a food preservative and veterinary antibiotic with more-stable alternatives. Lantibiotics appear to be widely distributed among sequenced *Geobacillus* species. For example, the genome of *G. kaustophilus* HTA426 contains two lantibiotic-biosynthesis gene clusters (centred on the genes for YP_146139 and YP_146147) that are both conserved in the recently sequenced *Geobacillus* sp. CAMR12739. The NG80-2 genome sequence also enabled discovery of the first nitrous oxide reductase gene from a Gram-positive, and a novel thermophilic long-chain alkane monooxygenase (Feng *et al.*, 2007). Furthermore, the genome sequence enabled proteomics-level confirmation of pathways for catabolism of long-chain alkanes (Feng *et al.*, 2007) and aromatics (Li *et al.*, 2012).

Many of the *Geobacillus* genome sequencing projects reported genes potentially encoding thermostable homologues of useful enzymes. In some cases, the genome sequences have been used to clone and express the genes of interest and characterize the enzyme for biotechnological potential. For example, the genome of *G. kaustophilus* HTA426 was recently mined for members of the glycoside hydrolase family 1, which have potential uses in synthesizing therapeutic oligosaccharides (Suzuki *et al.*, 2013). The genome sequence of the alkane-utilizing *G. thermoleovorans* B23 (Boonmak *et al.*, 2013) revealed a cluster of three long-chain alkane monooxygenase genes with homology to that of NG80-2 that showed activity *in vivo* when heterologously expressed in *Pseudomonas fluorescens* (Boonmak *et al.*, 2014). Recently, a novel thermostable endo-xylanase was cloned and expressed from *Geobacillus* sp. WSUCF1

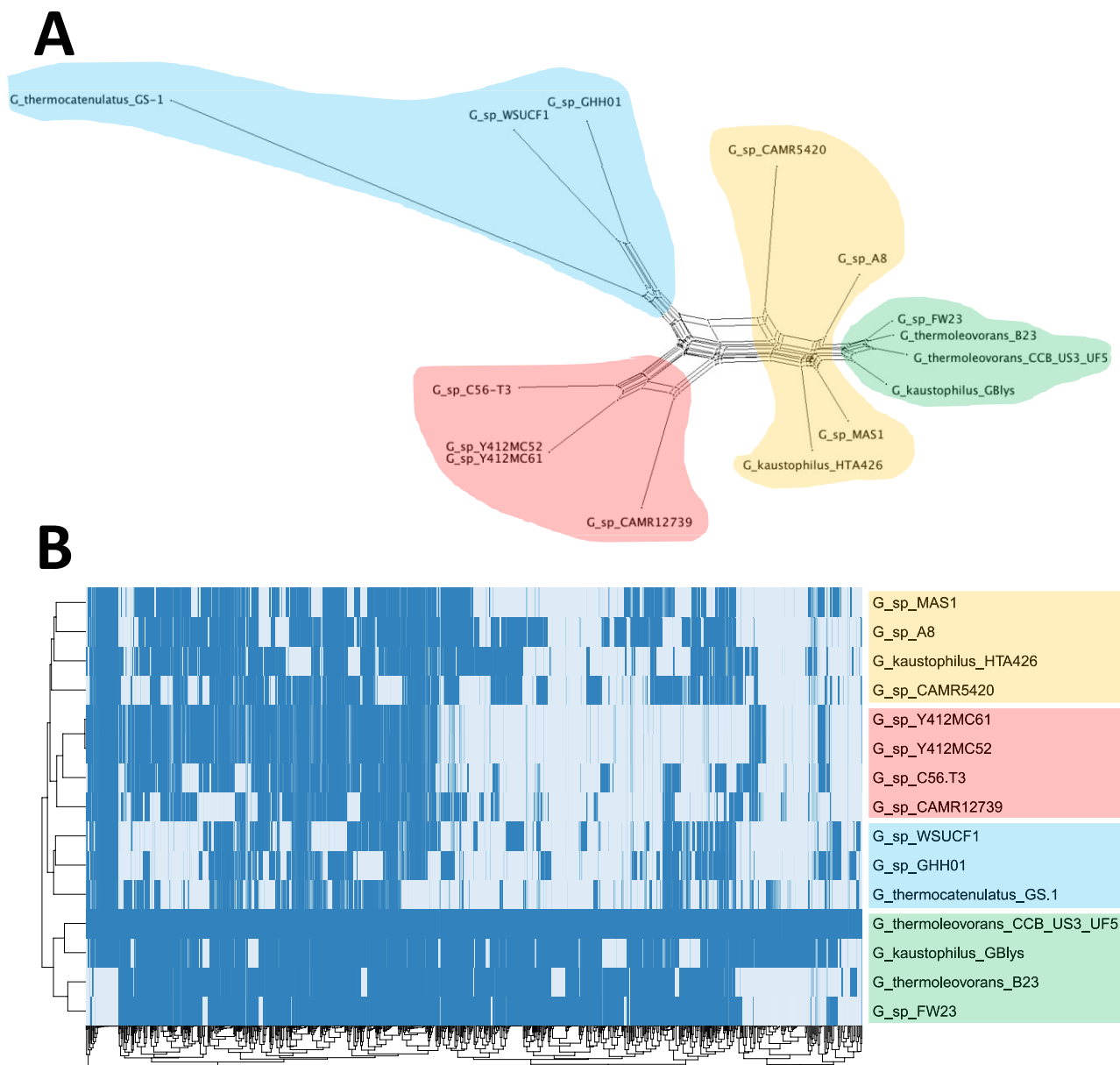


Fig. 3. Relationships among sequenced genomes within the *G. kaustophilus* clade resolved using whole-genome sequence data. The phylogenetic network in panel A was based on a concatenation of 1722 variant single-nucleotide sites in 1 874 967 nucleotides of the core genome present in all 15 genomes. The network was generated using the NEIGHBORNET algorithm (Bryant and Moulton, 2004) implemented in the SPLITSTREE software package (Huson, 1998). The heat-map in B indicates the presence (dark blue) and absence (light blue) of each of 931 non-core genes from the genome of *G. thermoleovorans* CCB US3 UF6 across the same 15 genomes appearing in A. The gene-content clusters are shaded in the same colours in both panels. The heat-map was rendered using Raivo Kolde's pheatmap package in R (R Development Core Team, R, 2013).

(Bhalla *et al.*, 2014) following the sequencing of its genome (Bhalla *et al.*, 2013).

Genome sequencing has revealed that interesting traits are often encoded on chromosomes rather than on the chromosome. For example, the biphenyl-degrading pathway of *Geobacillus* sp. JF8 (Mukerjee-Dhar *et al.*, 2005; Shintani *et al.*, 2014) and the long-chain alkane monooxygenase of *G. thermodenitrificans* NG80-2 (Feng

et al., 2007) are both located on plasmids. The dynamic loss and gain of such mobile elements presumably explains, in part, the physiological differences between natural isolates of *Geobacillus* spp. and it also suggests that these bacteria might be engineered to express new traits by introduction of recombinant plasmids. Indeed, progress has been made in developing plasmid shuttle vectors for heterologous expression in *Geobacillus* spp.

(Thompson *et al.*, 2008; Bartosiak-Jentys *et al.*, 2013).

The value of genome sequencing goes beyond cataloguing potentially useful enzymes, as exemplified by the recently published genomic study of strain NUB3621 (Blanchard *et al.*, 2014). Some previous attempts to fully exploit the potential of *Geobacillus* strains as whole-cell catalysts have been frustrated by the paucity of genetic and genomic resources (my own PhD research project in the mid-1990s being a case in point; Studholme, 1998). However, strain NUB3621 is a promising laboratory workhorse strain. It is one of the few *Geobacillus* strains that has been shown to be readily transformable with plasmid DNA (Wu and Welker, 1989); protocols have been developed for genetic analysis (Chen *et al.*, 1986) and a genetic map has been available for more than two decades (Vallier and Welker, 1990). Strain NUB3621 is a mutant derived from wild-type strain NUB36 that lacks its parent strain's restriction-modification system and this probably contributes to transformation efficiency. Incidentally, and consistent with this, we observed that transformation efficiency was significantly affected by the methylation status of the plasmid DNA (Thompson *et al.*, 2008).

Being one of the most genetically amenable *Geobacillus* strains, NUB3621 was obviously a high priority for genome sequencing. But rather than simply announcing and describing its genome sequence, the authors went on to show how the genome sequence could be exploited to further develop the strain as a host for heterologous expression and metabolic engineering (Blanchard *et al.*, 2014). Specifically, they used the genome sequence to clone two promoters and incorporated them into plasmid vectors: one for inducible gene expression and one constitutive. The authors also mention that they tried other promoters that did not work so well; presumably, the availability of the genome sequence allowed them to relatively quickly screen a number of candidates until they found the best ones. The combination of a genome sequence, allowing relatively facile construction of expression and/or knock-out constructs and a global view of metabolism, along with transformability and a wide range of growth temperatures [between 39 and 75°C (Wu and Welker, 1991)] make NUB3621 a strong candidate as the preferred thermophilic host for rationally designed metabolic engineering.

What's next?

The availability of complete (or nearly complete) genome sequences for nearly 30 *Geobacillus* strains (Table 1) as well as large-scale proteomic data for at least one (Feng *et al.*, 2007; Li *et al.*, 2012) should certainly accelerate cloning, expression and characterization of novel

thermostable and thermo-active enzymes, at least in an academic research context. However, there has been relatively little industrial uptake of enzymes from thermophiles, with much greater use of proteins originating from mesophiles but engineered for thermo-stability (Haki and Rakshit, 2003; Taylor *et al.*, 2011). The convergence of genomic data and transformability, at least for strain NUB3621, should help to remove the barriers to greater exploitation of thermophiles. However, genome sequences are not yet publicly available for the handful of other readily transformable *Geobacillus* strains such as *G. thermodenitrificans* K1041 (Narumi *et al.*, 1992), *G. stearothermophilus* IFO 12550 (Imanaka *et al.*, 1982), NRRL 1174 (Liao *et al.*, 1986) and *G. thermoglucosidasius* TN (Thompson *et al.*, 2008). Furthermore, although it is possible to predict the metabolic networks of bacteria from complete genome sequence, there is a need for comprehensive testing of these predictions through metabolomics. Only then can we rationally design genetic interventions to predictably manipulate metabolism. And finally, palaeo-genomics of ancient *Geobacillus* spores, which may be viable after billions of years of dormancy, might shed light on population-genetics and evolutionary processes over timescales that we previously assumed to be intractable (Nicholson, 2003; Zeigler, 2014).

Acknowledgements

The strains K1041, NUB3621 and NRRL1174 shown in Fig. 1 were kindly given by I. Narumi (Japan Atomic Energy Research Institute, Takasaki, Japan), N. Welker (Northwestern University, Evanston, USA) and H. Liao (Cangene, Ontario, Canada) respectively.

Conflict of interest

None declared.

References

- Banat, I.M., Marchant, R., and Rahman, T.J. (2004) *Geobacillus debilis* sp. nov., a novel obligately thermophilic bacterium isolated from a cool soil environment, and reassignment of *Bacillus pallidus* to *Geobacillus pallidus* comb. nov. *Int J Syst Evol Microbiol* **54**: 2197–2201.
- Bartosiak-Jentys, J., Hussein, A.H., Lewis, C.J., and Leak, D.J. (2013) Modular system for assessment of glycosyl hydrolase secretion in *Geobacillus thermoglucosidasius*. *Microbiology* **159**: 1267–1275.
- Begley, M., Cotter, P.D., Hill, C., and Ross, R.P. (2009) Identification of a novel two-peptide lantibiotic, lichenicidin, following rational genome mining for LanM proteins. *Appl Environ Microbiol* **75**: 5451–5460.
- Bhalla, A., Kainth, A.S., and Sani, R.K. (2013) Draft genome sequence of lignocellulose-degrading thermophilic

- bacterium *Geobacillus* sp. strain WSUCF1. *Genome Announc* **1**: pii: e00595-13. doi:10.1128/genomeA.00595-13.
- Bhalla, A., Bischoff, K.M., Uppugundla, N., Balan, V., and Sanj, R.K. (2014) Novel thermostable endo-xylanase cloned and expressed from bacterium *Geobacillus* sp. WSUCF1. *Bioresour Technol* **165**: 314–318.
- Blanchard, K., Robic, S., and Matsumura, I. (2014) Transformable facultative thermophile *Geobacillus stearothermophilus* NUB3621 as a host strain for metabolic engineering. *Appl Microbiol Biotechnol* **98**: 6715–6723.
- Boonmak, C., Takahashi, Y., and Morikawa, M. (2013) Draft genome sequence of *Geobacillus thermoleovorans* strain B23. *Genome Announc* **1**: pii: e00944-13. doi:10.1128/genomeA.00944-13.
- Boonmak, C., Takahashi, Y., and Morikawa, M. (2014) Cloning and expression of three *ladA*-type alkane monooxygenase genes from an extremely thermophilic alkane-degrading bacterium *Geobacillus thermoleovorans* B23. *Extremophiles* **18**: 515–523.
- Bryant, D., and Moulton, V. (2004) Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol* **21**: 255–265.
- Burgess, S.A., Lindsay, D., and Flint, S.H. (2010) Thermophilic bacilli and their importance in dairy processing. *Int J Food Microbiol* **144**: 215–225.
- Chen, X.-G., Stabnikova, O., Tay, J.-H., Wang, J.-Y., and Tay, S.T.-L. (2004) Thermoactive extracellular proteases of *Geobacillus caldoproteolyticus*, sp. nov., from sewage sludge. *Extremophiles* **8**: 489–498.
- Chen, Z.F., Wojcik, S.F., and Welker, N.E. (1986) Genetic analysis of *Bacillus stearothermophilus* by protoplast fusion. *J Bacteriol* **165**: 994–1001.
- Coorevits, A., Dinsdale, A.E., Halket, G., Lebbe, L., De Vos, P., Van Landschoot, A., and Logan, N.A. (2012) Taxonomic revision of the genus *Geobacillus*: emendation of *Geobacillus*, *G. stearothermophilus*, *G. jurassicus*, *G. toebii*, *G. thermodenitrificans* and *G. thermoglucosidans* (*nom. corrig.*, formerly ‘*thermoglucoisidans*’); transfer of *Bacillus thermantarcticus*. *Int J Syst Evol Microbiol* **62**: 1470–1485.
- Couñago, R., and Shamoo, Y. (2005) Gene replacement of adenylate kinase in the gram-positive thermophile *Geobacillus stearothermophilus* disrupts adenine nucleotide homeostasis and reduces cell viability. *Extremophiles* **9**: 135–144.
- Cripps, R.E., Eley, K., Leak, D.J., Rudd, B., Taylor, M., Todd, M., *et al.* (2009) Metabolic engineering of *Geobacillus thermoglucosidans* for high yield ethanol production. *Metab Eng* **11**: 398–408.
- De Maayer, P., Williamson, C.E., Vennard, C.T., Danson, M.J., and Cowan, D.A. (2014) Draft genome sequences of *Geobacillus* sp. strains CAMR5420 and CAMR12739. *Genome Announc* **2**: pii: e00567-14. doi:10.1128/genomeA.00567-14.
- Delcher, A.L., Phillippy, A., Carlton, J., and Salzberg, S.L. (2002) Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res* **30**: 2478–2483.
- Dinsdale, A.E., Halket, G., Coorevits, A., Van Landschoot, A., Busse, H.-J., De Vos, P., and Logan, N.A. (2011) Emended descriptions of *Geobacillus thermoleovorans* and *Geobacillus thermocatenulatus*. *Int J Syst Evol Microbiol* **61**: 1802–1810.
- Doi, K., Mori, K., Martono, H., Nagayoshi, Y., Fujino, Y., Tashiro, K., *et al.* (2013) Draft genome sequence of *Geobacillus kaustophilus* GBlys, a lysogenic strain with bacteriophage OH2. *Genome Announc* **1**: pii: e00634-13. doi:10.1128/genomeA.00634-13.
- Feng, L., Wang, W., Cheng, J., Ren, Y., Zhao, G., Gao, C., *et al.* (2007) Genome and proteome of long-chain alkane degrading *Geobacillus thermodenitrificans* NG80-2 isolated from a deep-subsurface oil reservoir. *Proc Natl Acad Sci USA* **104**: 5602–5607.
- Fridjonsson, O., Watzlawick, H., Gehweiler, A., and Mattes, R. (1999) Thermostable alpha-galactosidase from *Bacillus stearothermophilus* NUB3621: cloning, sequencing and characterization. *FEMS Microbiol Lett* **176**: 147–153.
- Garg, N., Tang, W., Goto, Y., Nair, S.K., and van der Donk, W.A. (2012) Lantibiotics from *Geobacillus thermodenitrificans*. *Proc Natl Acad Sci USA* **109**: 5241–5246.
- Haki, G.D., and Rakshit, S.K. (2003) Developments in industrially important thermostable enzymes: a review. *Bioresour Technol* **89**: 17–34.
- Huson, D.H. (1998) SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* **14**: 68–73.
- Imanaka, T., Fujii, M., Aramori, I., and Aiba, S. (1982) Transformation of *Bacillus stearothermophilus* with plasmid DNA and characterization of shuttle vector plasmids between *Bacillus stearothermophilus* and *Bacillus subtilis*. *J Bacteriol* **149**: 824–830.
- Li, Y., Wu, J., Wang, W., Ding, P., and Feng, L. (2012) Proteomics analysis of aromatic catabolic pathways in thermophilic *Geobacillus thermodenitrificans* NG80-2. *J Proteomics* **75**: 1201–1210.
- Liao, H., McKenzie, T., and Hageman, R. (1986) Isolation of a thermostable enzyme variant by cloning and selection in a thermophile. *Proc Natl Acad Sci USA* **83**: 576–580.
- Marchant, R., Sharkey, F.H., Banat, I.M., Rahman, T.J., and Perfumo, A. (2006) The degradation of n-hexadecane in soil by thermophilic geobacilli. *FEMS Microbiol Ecol* **56**: 44–54.
- Marks, T.J., and Hamilton, P.T. (2014) Characterization of a thermophilic bacteriophage of *Geobacillus kaustophilus*. *Arch Virol*. doi:10.1007/s00705-014-2101-8.
- Mead, D.A., Lucas, S., Copeland, A., Lapidus, A., Cheng, J.-F., Bruce, D.C., *et al.* (2012) Complete genome sequence of *Paenibacillus* strain Y4.12MC10, a novel *Paenibacillus lautus* strain isolated from Obsidian hot spring in Yellowstone National Park. *Stand Genomic Sci* **6**: 381–400.
- Muhd Sakaff, M.K.L., Abdul Rahman, A.Y., Saito, J.A., Hou, S., and Alam, M. (2012) Complete genome sequence of the thermophilic bacterium *Geobacillus thermoleovorans* CCB_US3_UF5. *J Bacteriol* **194**: 1239.
- Mukerjee-Dhar, G., Shimura, M., Miyazawa, D., Kimbara, K., and Hatta, T. (2005) *bph* genes of the thermophilic PCB degrader, *Bacillus* sp. JF8: characterization of the divergent ring-hydroxylating dioxygenase and hydrolase genes

- upstream of the Mn-dependent BphC. *Microbiology* **151**: 4139–4151.
- Narumi, I., Sawakami, K., Nakamoto, S., Nakayama, N., Yanagisawa, T., Takahashi, N., and Kihara, H. (1992) A newly isolated *Bacillus stearothermophilus* K1041 and its transformation by electroporation. *Biotechnol Tech* **6**: 83–86.
- Nazina, T.N., Tourova, T.P., Poltarau, A.B., Novikova, E., V., Grigoryan, A.A., Ivanova, A.E., *et al.* (2001) Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzonensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustophilus*, *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus*, *G. th.* *Int J Syst Evol Microbiol* **51**: 433–446.
- Nicholson, W.L. (2003) Using thermal inactivation kinetics to calculate the probability of extreme spore longevity: implications for paleomicrobiology and lithopanspermia. *Orig Life Evol Biosph* **33**: 621–631.
- Niehaus, F., Bertoldo, C., Kähler, M., and Antranikian, G. (1999) Extremophiles as a source of novel enzymes for industrial application. *Appl Microbiol Biotechnol* **51**: 711–729.
- Pore, S.D., Arora, P., and Dhakephalkar, P.K. (2014) Draft genome sequence of *Geobacillus* sp. strain FW23, isolated from a formation water sample. *Genome Announc* **2**: pii: e00352-14. doi:10.1128/genomeA.00352-14.
- R Development Core Team, R. (2013) R: a language and environment for statistical computing. *R Found Stat Comput* **1**: 409.
- Sandalli, C., Singh, K., Modak, M.J., Ketkar, A., Canakci, S., Demir, I., and Belduz, A.O. (2009) A new DNA polymerase I from *Geobacillus caldoxylosilyticus* TK4: cloning, characterization, and mutational analysis of two aromatic residues. *Appl Microbiol Biotechnol* **84**: 105–117.
- Schmidt-Dannert, C., Pleiss, J., and Schmid, R.D. (1998) A toolbox of recombinant lipases for industrial applications. *Ann N Y Acad Sci* **864**: 14–22.
- Seo, M.-J., Lee, B.-S., Pyun, Y.-R., and Park, H. (2011) Isolation and characterization of N-acylhomoserine lactonase from the thermophilic bacterium, *Geobacillus caldoxylosilyticus* YS-8. *Biosci Biotechnol Biochem* **75**: 1789–1795.
- Shintani, M., Ohtsubo, Y., Fukuda, K., Hosoyama, A., Ohji, S., Yamazoe, A., *et al.* (2014) Complete genome sequence of the thermophilic polychlorinated biphenyl degrader *Geobacillus* sp. strain JF8 (NBRC 109937). *Genome Announc* **2**: e01213–13.
- Siddiqui, M.A., Rashid, N., Ayyampalayam, S., and Whitman, W.B. (2014) Draft genome sequence of *Geobacillus thermopakistanensis* strain MAS1. *Genome Announc* **2**: pii: e00559-14. doi:10.1128/genomeA.00559-14.
- Studholme, D.J. (1998) Metabolic engineering of thermophilic *Bacillus* species for ethanol production. PhD Thesis. London: Department of Biochemistry, Imperial College.
- Studholme, D.J., Jackson, R.A., and Leak, D.J. (1999) Phylogenetic analysis of transformable strains of thermophilic *Bacillus* species. *FEMS Microbiol Lett* **172**: 85–90.
- Suzuki, H., Okazaki, F., Kondo, A., and Yoshida, K. (2013) Genome mining and motif modifications of glycoside hydrolase family 1 members encoded by *Geobacillus kaustophilus* HTA426 provide thermostable 6-phospho- β -glycosidase and β -fucosidase. *Appl Microbiol Biotechnol* **97**: 2929–2938.
- Tabachnikov, O., and Shoham, Y. (2013) Functional characterization of the galactan utilization system of *Geobacillus stearothermophilus*. *FEBS J* **280**: 950–964.
- Takami, H., Takaki, Y., Chee, G.-J., Nishi, S., Shimamura, S., Suzuki, H., *et al.* (2004) Thermoadaptation trait revealed by the genome sequence of thermophilic *Geobacillus kaustophilus*. *Nucleic Acids Res* **32**: 6292–6303.
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**: 2725–2729.
- Taylor, M.P., Eley, K.L., Martin, S., Tuffin, M.I., Burton, S.G., and Cowan, D.A. (2009) Thermophilic ethanologenes: future prospects for second-generation bioethanol production. *Trends Biotechnol* **27**: 398–405.
- Taylor, M.P., Zyl, L., van Tuffin, I.M., Leak, D.J., and Cowan, D.A. (2011) Genetic tool development underpins recent advances in thermophilic whole-cell biocatalysts. *Microb Biotechnol* **4**: 438–448.
- Thompson, A.H., Studholme, D.J., Green, E.M., and Leak, D.J. (2008) Heterologous expression of pyruvate decarboxylase in *Geobacillus thermoglucosidasius*. *Biotechnol Lett* **30**: 1359–1365.
- Vallier, H., and Welker, N.E. (1990) Genetic map of the *Bacillus stearothermophilus* NUB36 chromosome. *J Bacteriol* **172**: 793–801.
- Wiegand, S., Rabausch, U., Chow, J., Daniel, R., Streit, W.R., and Liesegang, H. (2013) Complete genome sequence of *Geobacillus* sp. strain GHH01, a thermophilic lipase-secreting bacterium. *Genome Announc* **1**: e0009213. doi:10.1128/genomeA.00092-13.
- Wiegel, J., Ljungdahl, L.G., and Demain, A.L. (1985) The importance of thermophilic bacteria in biotechnology. *Crit Rev Biotechnol* **3**: 39–108.
- Wu, L., and Welker, N.E. (1991) Temperature-induced prote in synthesis in *Bacillus stearothermophilus* NUB36. *J Bacteriol* **173**: 4889–4892.
- Wu, L.J., and Welker, N.E. (1989) Protoplast transformation of *Bacillus stearothermophilus* NUB36 by plasmid DNA. *J Gen Microbiol* **135**: 1315–1324.
- Xu, K., Su, F., Tao, F., Li, C., Ni, J., and Xu, P. (2013) Genome sequences of two morphologically distinct and thermophilic *Bacillus coagulans* strains, H-1 and XZL9. *Genome Announc* **1**: 4563–4564.
- Yao, N., Ren, Y., and Wang, W. (2013) Genome sequence of a thermophilic *Bacillus*, *Geobacillus thermodenitrificans* DSM465. *Genome Announc* **1**: pii: e01046-13. doi:10.1128/genomeA.01046-13.
- Zeigler, D.R. (2005) Application of a recN sequence similarity analysis to the identification of species within the bacterial genus *Geobacillus*. *Int J Syst Evol Microbiol* **55**: 1171–1179.

Zeigler, D.R. (2014) The *Geobacillus* paradox: why is a thermophilic bacterial genus so prevalent on a mesophilic planet? *Microbiology* **160**: 1–11.

Zhao, Y., Caspers, M.P., Abee, T., Siezen, R.J., and Kort, R. (2012) Complete genome sequence of *Geobacillus*

thermoglucosidans TNO-09.020, a thermophilic spore-former associated with a dairy-processing environment. *J Bacteriol* **194**: 4118.