The Geobacillus plasmid set: a modular toolkit for thermophile engineering

Benjamin Reeve, Elena Martinez-Klimova, Joachim de Jonghe, David J. Leak and Tom Ellis*

Centre for Synthetic Biology and Innovation, Department of Bioengineering Imperial College London, London SW7 2AZ, UK.

Keywords: Thermophile, Plasmid, Geobacillus, Genetic tools

Supporting Information Placeholder

ABSTRACT: Geobacillus thermoglucosidasius is a gram-positive thermophile of industrial interest that exhibits rapid growth and can utilize a variety of plant-derived feedstocks. It is an attractive chassis organism for high temperature biotechnology and synthetic biology applications but is currently limited by a lack of available genetic tools. Here we describe a set of modular shuttle vectors, including a promoter library and reporter proteins. The compact plasmids are composed of interchangeable modules for molecular cloning in Escherichia coli and stable propagation in G. thermoglucosidasius and other Geobacillus species. Modules include two origins of replication, two selectable markers and three reporter proteins for characterization of gene expression. For fine-tuning heterologous expression from these plasmids, we include a characterized promoter library and test ribosome binding site design. Together, these gene expression tools and a standardized plasmid set can facilitate modularity and part exchange to make Geobacillus a thermophile chassis for synthetic biology.

Synthetic biology relies on characterized parts and tools to build new functionalities into organisms. The availability and ease of exchange of these parts limits chassis choice and hence most research takes place in model organisms such as E. coli or S. cerevisiae. Whilst recent work has expanded synthetic biology research and application into many non-model microorganisms, none of these grow at temperatures outside the normal mesophilic range (10 to 40 °C). Thermophiles are a major class of microorganisms of particular value for industrial applications and thermophilic bacteria are under-utilized considering their many potential advantages as chassis organisms. They benefit from rapid growth and feedstock conversion plus metabolic versatility particularly in the utilization of plant biomass. Risk of contamination during large-scale growth is reduced, as common contaminant species are mesophiles, unable to grow at higher temperatures. Finally, high temperature fermentations can reduce the need for additional cooling or heating after feedstock pretreatments or before product recovery.

Geobacillus thermoglucosidasius (alternatively called thermoglucosidans) is a gram-positive endospore-forming, facultatively anaerobic thermophilic bacterium. In nature, Geobacillus species have been isolated from diverse warm habitats including volcanic springs, hydrothermal vents, oil wells and compost heaps. Members of this genus exhibit growth over a wide range of temperatures, growing well between 40 and 70 °C with optimal growth at 55 to 65 °C. G. thermoglucosidasius is a particularly attractive chassis for industrial biotechnology able to utilize a broad range of feedstocks, uptaking oligomeric carbohydrates and metabolizing both hexose and pentose sugars found in lignocellulosic biomass. It is a promising chassis organism for commodity chemical production and for bioremediation. Here we present a genetic toolkit for engineering G. thermoglucosidasius, with components assembled in modular combinations to create standardized plasmids with greater versatility and transformation efficiency than existing vectors. The plasmid set includes origins of replication, selectable markers and a choice of three reporter genes expressed from a novel promoter library.
Previous engineering in *G. thermoglucosidasius* has been limited to gene knockouts or overexpression from natural promoters\(^{10,12}\). Libraries of Promoters and ribosome binding sites however, allow the more precise tuning of gene expression demanded by modern synthetic biology and metabolic engineering. The strongest and most commonly used promoter in *G. thermoglucosidasius* to date is the *G. stearothermophilus* lactate dehydrogenase promoter pLdh\(^{10,13}\), however expression levels are influenced by redox conditions, so this may not be universally useful\(^{14}\). The toolkit here includes a stronger and more constitutive promoter, pRplS from which a promoter library has been generated. We also demonstrate that a synthetic RBS library can be designed with existing tools\(^{15}\) to further tune protein translation rate.

The architecture of plasmids in the toolkit is illustrated in Figure 1, with the naming convention in Table 1. The interchangeable modules are separated by rare-cutter restriction sites in a format inspired by previously described Clostridia pMTL\(^{16}\) and SEVA plasmid collections\(^{17}\). Pre-existing shuttle vectors described for *G. thermoglucosidasius* include pUCG18,\(^{13}\) pUCG3.8,\(^{18}\) pBST22\(^{19}\) and pUB190.\(^{10}\) With the exception of pUCG3.8, these are large plasmids based on naturally occurring elements, and all have comparatively little characterization. The Geobacillus plasmid set provides improved, modular, characterized plasmids with increased electroporation efficiencies (Table 2).

For propagation in *E. coli* during plasmid cloning and construction, a ColE1 origin of replication module is included in all plasmids.\(^{20}\)

---

**Figure 1.** The *Geobacillus* plasmid set architecture. (a) Diagram of the 5-part plasmid including two antibiotic resistance markers, the pRplS promoter plus a selection of reporter genes (b) Diagram of the 4-part plasmid including synthetic Multiple Cloning Site.

**Table 1. Names and contents of Geobacillus plasmid set plasmids** All plasmids follow the convention *pGxxx-cargo* where the first variable is the replicon, repBST1 = 1, repB = 2, followed by letters indicating the selection marker(s) present and then the cargo. The above plasmids are available from AddGene with sequences in the NCBI database, accession numbers in Supporting Table 1. Alternative module combinations can be created through simple restriction cloning.

For propagation in *G. thermoglucosidasius* two replication origins are included. First, repBST1 originating from pBST22\(^{19}\) which was derived from *Bacillus stearothermophilus* cryptic plasmid pBST1. This replicon has previously been used in the *G. thermoglucosidasius* plasmid pUCG18.\(^{13}\) Secondly repB, a temperature sensitive replicon, inactive over 68 °C, which is useful for the creation of knock-out strains.\(^{10}\) This sequence was obtained from pUB110\(^{42}\) and originates from a cryptic *Staphylococcus aureus* plasmid.
Both replicons are functional in *G. thermoglucosidasius* between 45-65 °C with copy numbers per chromosome estimated by quantitative PCR (qPCR) to be approximately 160 for repBSTI and 115 for repB at 55 °C (Figure 3a). Both replicons allow plasmids to be stably maintained with antibiotic selection for at least 5 days of continuous growth where dilution into fresh media is performed every 12 hours (Supporting Figure 1). These replication origins were not found to be compatible. Alternative replicons would therefore be required in order to propagate two plasmids in one cell simultaneously and would be a priority for further expansion of this toolkit.

For selection in *E. coli* the commonly used *bla* gene for ampicillin resistance (*ampR*) is included as this marker was found to give the highest transformation efficiencies in *E. coli* (Table 2). For selection in both *G. thermoglucosidasius* and *E. coli*, two options are provided. First, TK101: a thermostable kanamycin resistance gene (kanR). TK101 was obtained from pBST22 and is a variant of the mesophilic gene found in pUB110. Previous vectors employing this marker lacked a transcriptional terminator after the gene; this was corrected for the plasmids presented here. The second marker, CatE, is from the *Staphylococcus aureus* plasmid pC194. As described by Kobayashi et al. a single base substitution (G to A) at base 412 was made via site directed mutagenesis to improve thermostability by changing an alanine to threonine amino acid (A138T). This probably improves hydrogen bonding interactions in the protein backbone, increasing stability without compromising catalytic activity. Kanamycin and chloramphenicol were chosen for selection because these compounds have good thermostability compared to other commonly used antibiotics and have previously been used for selection in thermophiles. The kanamycin resistance marker in pG1K gives significantly better transformation efficiency than the chloramphenicol marker in pG1C, likely due to higher thermostability, and so we recommend pG1K as the first choice backbone. The Geobacillus plasmid set includes two architectures, 4-part and 5-part. In addition to cargo modules and separate origins of replication for each species, the 5-part plasmids include the *bla* ampicillin resistance gene for *E. coli*. The 4-part plasmids omit this to reduce plasmid size and increase transformation efficiency in *G. thermoglucosidasius*. As electroporation efficiency is negatively correlated with plasmid size, compact vector backbones increase efficiency and allow larger cargoes such as multi-gene operons to be carried whilst maintaining workable efficiency. A novel Multiple Cloning Site (MCS) was designed containing commonly used restriction sites (Supporting Figure 2). Two transcriptional terminators, rho1 and rho2, insulate this module; both have previously been used in constructs for *Geobacillus* species. Terminator rho1 is taken from pUCG18 and rho2 from plasmids used for secretion in Geobacillus. They resemble typical rho-independent terminators consisting of a stable hairpin structure followed by polyT.

Transformation efficiency in both *E. coli* and *G. thermoglucosidasius* for the Geobacillus plasmid set was compared with previous vectors (Table 2). The new vectors all show good transformation efficiencies, with pG1K giving over an order of magnitude improvement compared to the best previously published vector pUCG3.8. In addition to the new plasmid's more compact size, the improvement in transformation efficiency could also be due to improved TK101 kanR expression, as a transcriptional terminator was added here that was previously absent. pG1K was also found to transform other *Geobacillus* species, *G. kaustophilus*, *G. thermoleovorans* and *G. thermodenitrificans* with efficiencies between $10^3$ and $10^5$ CFU/µg DNA (Supporting Table 2).

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Size/kbp</th>
<th>Antibiotic selection</th>
<th><em>G. thermoglucosidasius</em></th>
<th><em>E. coli</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>pUCG18</td>
<td>6.3</td>
<td>Kanamycin (Ampicillin)</td>
<td>4.9×10^3 ±22%</td>
<td>1.6×10^6 ±7% (4.4×10^6 ±19%)</td>
</tr>
<tr>
<td>pUCG3.8</td>
<td>3.8</td>
<td>Kanamycin</td>
<td>5.2×10^3 ±27%</td>
<td>1.9×10^6 ±6%</td>
</tr>
<tr>
<td>pG1K</td>
<td>3.7</td>
<td>Kanamycin</td>
<td>5.3×10^4 ±23%</td>
<td>3.4×10^6 ±7%</td>
</tr>
<tr>
<td>pG2K</td>
<td>3.8</td>
<td>Kanamycin</td>
<td>1.1×10^5 ±12%</td>
<td>3.5×10^6 ±2%</td>
</tr>
<tr>
<td>pG1C</td>
<td>3.9</td>
<td>Chloramphenicol</td>
<td>3.9×10^3 ±32%</td>
<td>9.6×10^4 ±42%</td>
</tr>
<tr>
<td>pG1AK</td>
<td>4.7</td>
<td>Kanamycin (Ampicillin)</td>
<td>5.8×10^3 ±49%</td>
<td>3.0×10^6 ±9% (7.4×10^6 ±9%)</td>
</tr>
</tbody>
</table>

Table 2. Transformation efficiencies of Geobacillus plasmid set plasmids compared to the most efficient existing vectors, pUCG18 and pUCG3.8. Percent error shown is the standard deviation of three biological replicates. pG1K transformation efficiency is significantly greater (P=0.019) than both pUCG18 and pUCG3.8.
Three reporter proteins, functional in both *G. thermoglucosidasius* and *E. coli*, are provided (Figure 2, with *E. coli* data in Supporting Figure S3). Superfolder green fluorescent protein, sfGFP\(^\text{2}^\text{7}\), shows strong brightness and thermostability, maintaining fluorescence up to 70 °C in vivo (26). As an alternative fluorescent reporter, mCherry\(^\text{2}^\text{8}\) is also included, however this is less thermostable and loses fluorescence at growth temperatures over 50 °C (Figure 2a, b). The third reporter protein is the previously described enzyme PheB\(^\text{1}^\text{4}\) taken from *G. stearothermophilus* which reports gene expression via a colorimetric assay after the addition of catechol reagent (Figure 2c). The PheB enzyme does not require oxygen to fold so may be used as a reporter under anaerobic growth conditions. An anaerobic fluorescent reporter would be preferred, however two thermostable variants of flavin-binding LOV-domain fluorescent proteins that were tested not found to be functional in *G. thermoglucosidasius* (see Supporting Table 3 and Supporting Methods).

The *G. thermoglucosidasius* promoter pRps exhibits very strong and constitutive expression in both *E. coli* and *G. thermoglucosidasius*. The wild type promoter sequence was taken from the *G. thermoglucosidasius* genome sequence (28) where it promotes transcription of the ribosomal protein RplS, observed in other *Bacilli* to be strongly expressed in all growth conditions.\(^\text{3}^\text{1}\) The previously used pLdh promoter, included for comparison (sequence in Supporting Table 4), is not as constitutive\(^\text{1}^\text{8}\) and does not give expression in *E. coli*. To construct a fine-graded constitutive promoter library for *G. thermoglucosidasius*, mutagenic PCR with synthetic nucleotide analogs was used to amplify the pRpsWT sequence, incorporating mutations at a rate of approximately 10%. The amplified mutated promoter region was assembled with a strong RBS sequence from the PheB gene of *G. stearothermophilus* and expression levels of approximately 100 clones were assayed using the sfGFP reporter as an output. Twenty mutant clones covering a range of expression levels were selected to form the library. These were sequenced through the promoter region (Supporting Table 5) and characterized by sfGFP expression in both *E. coli* and *G. thermoglucosidasius* (Figure 3b).

The library spans a 100-fold range in expression strength and includes members that are significantly stronger than pLdh. The library also includes promoter pairs with similar strengths (for example Rps1 & WT or S & 6) but significantly different promoter sequences. A comparison of expression levels between the two species (Figure 3c) for the library promoters shows a very low correlation (\(R^2 = 0.042\)), with some library members that are weak in *E. coli* being strong in *G. thermoglucosidasius* and vice versa. These differences between expression levels can be of benefit. For example, a set of promoters with low strength in *E. coli* but covering the full range of expression in *G. thermoglucosidasius* can be selected from this library. This subset could be useful for cloning certain genes as low *E. coli* expression will reduces stresses caused to this host and therefore should improve efficiency of the cloning and propagation stages in *E. coli*.
Figure 3. (a) Plasmid copy number per chromosome estimated by qPCR for plasmids pG1AK and pG2AK with the two different Geobacilli replicons, repBSTI and repB respectively at different growth temperatures. Error bars show standard deviation of three biological repeats (b) Fluorescence outputs of Rpl promoter library members in G. thermoglucosidasius above, and E. coli below, the axis. The strong G.st RBS sequence is used in all cases (c) Correlation between promoter outputs in the two species. A very weak positive correlation of $R^2 = 0.042$ is seen. (d) Ribosome binding site library characterization in the two species. All GFP fluorescence readings were corrected for media autofluorescence and divided by corrected OD600 readings, error bars show standard deviation of three biological replicates.

To allow further tuning of gene expression, we investigated whether reporter output could be predictably modified by altering ribosome binding site (RBS) sequence. As the efficiency of RBS sequences is highly dependent on the context of their local sequence, it is important that new RBS parts can be reliably designed. A small 4-member library was designed using the Salis Lab Ribosome Binding Site Calculator software tool15 (sequences in Supporting Table 6). Parameters for design were set to give a wide dynamic range in expression with RBS-A designed using the maximum translational initiation rate and RBS-D set to give low translation initiation. The RBS library sequences were characterized between the pRplSWT promoter and sfGFP reporter gene in plasmid pG1AK for both G. thermoglucosidasius and E. coli. Expression output is compared to the RBS sequence from the previously reported G. stearothermophilus (G.st) PheB gene14 (Figure 3d). For E. coli the predicted rank order of RBS strength is correct including the natural G.st RBS. In G. thermoglucosidasius the natural G.st RBS sequence is stronger than all designed RBS sequences. For the synthetic sequences however, despite the RBS calculator being intended for predicting translation rate in mesophilic organisms, the rank order of expression strengths matched the intended design. The rank of the natural G.st RBS was not correctly predicted possibly due to differences in Gram-positive translation. For the designed sequences however the RBS Calculator was able to predictably vary translation initiation rates in the Gram-positive Geobacillus thermophile and so could be a useful tool for synthetic biology in this organism.

G. thermoglucosidasius is a promising chassis for thermophile synthetic biology but its use has been limited by availability of versatile genetic tools. The plasmids and part libraries presented here will make future engineering of Geobacilli more rational and reliable and aid characterization of further new genetic parts. The plasmids are available from Addgene with sequences in the NCBI database (accession numbers in Supporting Table 1).
Materials and Methods

Mutagenic PCR The reaction comprised 10x standard Taq Mg- free reaction buffer (NEB UK), 50 mM MgCl2, 200 μM of dPTP and 8-oxo-dGTP, 1 mg/ml gelatine and 5 U/μl Taq polymerase. 8-Oxo-dGTP can mispair with A, leading to A-to-C and G-to-T transition mutations. dPTP in combination with 8-Oxo-dGTP can cause both transition mutations (A-to-G and G-to-A) and transversion mutations (A-to-C and G-to-T). PCR settings were: 98 °C for 2 minutes, then cycles of 56 °C for 1 minute (denaturation), 55 °C for 1.5 minutes (annealing), 72 °C for 5 minutes (elongation), for 20 cycles giving a mutation rate of approximately 10% and finally 5 minutes of extension at 72 °C. The mixes were then treated with 0.5 μl DpnI (NEB UK) at 37 °C for 1 hour to digest template DNA and used as template for a further PCR amplification with phusion polymerase to increase concentration and add overlap sequences for Gibson assembly into the backbone vector.

Estimation of transformation efficiency All Geobacillus strains were transformed via an adapted high osmolarity electroporation method previously shown to be successful in Geobacillus. Chemically competent E. coli was transformed by a heat shock method (details of strains in Supporting Table 7 and methods in Supporting Methods). Transformation efficiencies are displayed in colony forming units (CFU) per microgram of DNA. Efficiencies were estimated from the average number of colonies produced from three electroporations with 30 ng of DNA. After recovery the cells were serially diluted and plated onto agar plates with the appropriate antibiotics. Colonies were counted after 24 hours incubation at 37 °C for E. coli and 55 °C for Geobacillus species. P values for statistical significance were calculated by independent two-sample t-tests.

Site directed mutagenesis To mutate a single basepair in CatE phosphorylated primers were ordered (IDT Inc.) including the mutated base on the end of one primer. Plasmid pGIAC was amplified via PCR as above with the phosphorylated primers then template DNA was digested with DpnI restriction enzyme (NEB UK). The PCR product was self-ligated using T4 ligase (NEB UK) according to manufacturers instructions and transformed into E. coli with selection on chloramphenicol supplemented LBagar plates.

Parts characterization with sfGFP For G. thermoglucosidasius cultures were grown from single colonies in 5 ml of 2SPYNG media in 50 ml tubes at 55 °C overnight with shaking at 200 rpm. Cultures were then diluted 100x into fresh media and grown to stationary phase (maximum OD600). For each tube, a 200 μl aliquot of culture was added to a clear, flat bottom 96-well plate and GFP plus OD600 measurements were made with a BioTek Synergy HT plate reader (Biotek Inc. USA). For E. coli, LB media at 37 °C was used and cells were grown in 200 μl of media in 96-well plates. After subtracting for media auto fluorescence GFP readings were divided by OD600 readings to give an estimate of GFP per cell.

Plasmid copy number estimation Per chromosome estimates of plasmid copy number were determined by qPCR as described by Lee et al. 2006 and Skulj et al 2008. Details are given in the supporting methods.

ASSOCIATED CONTENT

Supporting information is available on the ACS publications website.

Figure S1. Plasmid segregational stability, S2. Multiple cloning site design, S3. Reporter gene function in E. coli Tables S1. Plasmid accession numbers for AddGene and NCBI, S2. Transformation efficiencies of other Geobacillus species, S3. Sequence data of Flavin-based fluorescent proteins considered for expression in G. thermoglucosidasius. S4. pLdh promoter sequence, S5. pRpsL promoter library sequences, S6. RBS sequences, S7. Bacterial strains used in this study

Supporting methods and references

AUTHOR INFORMATION

Corresponding Author
*E-mail: t.ellis@imperial.ac.uk

Author Contributions
B.R. and E.M.K constructed and tested the plasmids, analyzed data and created figures. J.J. assisted in constructing and analyzing the promoter library. B.R., E.M.K and T.E. wrote the manuscript. T.E. and D.J.L designed the project, supervised and coordinated the research.

Notes
The authors declare no competing financial interest.

Funding Sources
This work was funded by EPSRC awards EP/G036004/1 and EP/J02175X/1. B.R. was funded by the Department of Bioengineering, Imperial College London. E.M.K. was funded by CONACyT-México. D.J.L. is funded by BBSRC awards BB/J019445/1, BB/J001120/2 and EPSRC awards EP/K014889/1, EP/K014897/1.

ACKNOWLEDGEMENT

The authors would like to thank Jeremy Bartiosak-Jentys and Ali Hussein from the Leak lab, University of Bath for advice and materials. They would also like to thank undergraduate project students Bianca de Blasi, Henrik Hagemann, Joce- lyne Holdbrook, Sophie Morse and Yutong Wu who helped with Geobacillus work.
REFERENCES

Modular plasmids for thermophilic bacteria chassis

70°C

40°C

Characterized promoters & reporter proteins