Greedy reduction of Bacillus subtilis genome yields emergent phenotypes of high resistance to a DNA damaging agent and low evolvability

Etienne Dervyn, Anne-Gaëlle Planson, Kosei Tanaka, Victor Chubukov, Cyprien Guérin, Sandra Derozier, François Lecointe, Uwe Sauer, Ken-Ichi Yoshida, Pierre Nicolas, et al.

2974–2992 *Nucleic Acids Research, 2023, Vol. 51, No. 6* *Published online 15 March 2023*
https://doi.org/10.1093/nar/gkad145

Greedy reduction of *Bacillus subtilis* genome yields emergent phenotypes of high resistance to a DNA damaging agent and low evolvability

Etienne Dervyn [1,†], Anne-Gaëlle Planson [1,†], Kosei Tanaka[1,†], Victor Chubukov[2], Cyprien Guérin [3], Sandra Derozier[3], François Lecointe[1], Uwe Sauer[2], Ken-Ichi Yoshida[4], Pierre Nicolas[3], Philippe Noirot[1,*] and Matthieu Jules [1,*]

[1]Université Paris-Saclay, INRAE, AgroParisTech, Micalis Institute, 78350 Jouy-en-Josas, France, [2]Institute of Molecular Systems Biology, ETH Zurich, 8093 Zurich, Switzerland, [3]Université Paris-Saclay, INRAE, MaIAGE, 78350 Jouy-en-Josas, France and [4]Graduate School of Science, Technology and Innovation, Kobe University, Kobe, Japan

Received July 04, 2022; Revised February 09, 2023; Editorial Decision February 11, 2023; Accepted February 17, 2023

ABSTRACT

Genome-scale engineering enables rational removal of dispensable genes in chassis genomes. Deviating from this approach, we applied greedy accumulation of deletions of large dispensable regions in the *Bacillus subtilis* genome, yielding a library of 298 strains with genomes reduced up to 1.48 Mb in size. High-throughput physiological phenotyping of these strains confirmed that genome reduction is associated with substantial loss of cell fitness and accumulation of synthetic-sick interactions. Transcriptome analysis indicated that <15% of the genes conserved in our genome-reduced strains exhibited a twofold or higher differential expression and revealed a thiol-oxidative stress response. Most transcriptional changes can be explained by loss of known functions and by aberrant transcription at deletion boundaries. Genome-reduced strains exhibited striking new phenotypes relative to wild type, including a very high resistance (increased >300-fold) to the DNA-damaging agent mitomycin C and a very low spontaneous mutagenesis (reduced 100-fold). Adaptive laboratory evolution failed to restore cell fitness, except when coupled with a synthetic increase of the mutation rate, confirming low evolvability. Although mechanisms underlying this emergent phenotype are not understood, we propose that low evolvability can be leveraged in an engineering strategy coupling reductive cycles with evolutive cycles under induced mutagenesis.

INTRODUCTION

Genome engineering methods were developed to build streamlined host strains (i.e. chassis) suitable for industrial applications. As methods gained in efficiency, engineering at genome-scale also served to investigate fundamental evolutionary questions (1). Genome editing and recoding was set to accelerate the creation of genetically engineered microbes and allowed the expansion of the genetic code (2–4). Genome synthesis led to a first experimentally tested minimal set of genes required to sustain life (5,6). Genome reduction by iterative deletion of precisely defined dispensable regions identified novel essential and co-essential functions (7–11). Although genome streamlining using large-scale genome engineering is limited to model organisms, it revealed general properties of genome-reduced microorganisms such as the loss of cell fitness (3,11–13) and genetic competence (12). Genomic regions harboring horizontally transferred genes were found to be important for cell fitness because their deletion triggers a general stress response (14). Finally, when reducing the *Escherichia coli* genome, spontaneous mutation rates have been reported to increase with deletion size, suggesting a negative correlation with genome size (15).

Natural evolution allows for genome complexification and streamlining. Microbes are considered to be near an optimal adaptive fitness corresponding either to a specific ecological niche or to a wide range of environmental growth conditions. Evolution of prokaryotic genomes is dominated by long reductive phases by loss of genetic material, interspaced with shorter, explosive phases of genome complexification by mutation and horizontal gene transfer. Laboratory evolution of genome-reduced strains can help correct

*To whom correspondence should be addressed. Tel: +33 1 34 65 29 56; Email: matthieu.jules@inrae.fr
Correspondence may also be addressed to Philippe Noirot. Email: philippe.noirot@inrae.fr
†The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

growth defects as natural evolution does and may help decipher evolutionary processes. For instance, adaptive laboratory evolution of an *E. coli* strain with a genome reduced by 23% revealed that growth recovery is due to the correction of metabolic imbalances associated with transcriptome- and proteome-wide remodeling, which could not be predicted based on current knowledge (13). Also, in genome-reduced strains of *E. coli*, adaptive evolution by serial transfer of the genome-reduced strains improved growth rate and reduced mutation rates to wild type level (15). It is known that mutation rate is not constant as it can vary periodically with the cell cycle (16) and locally within genotypes in response to various stress, ageing or population density (17,18), although the interpretation of stress-induced mutagenesis as an adaptive strategy has been questioned and might also reflect selective enrichment of mutants in ageing populations (19). Adaptive laboratory evolution experiments however are limited to a narrow range of growth environments and the inherent genetic content of the organism under study, which limits genome evolution (20,21).

We investigate the consequences of genome reduction in *Bacillus subtilis,* which has many advantages for genome engineering including natural competence, efficient chromosomal replacement, existing repertoires of non-essential regions and genes (10,22,23), and availability of libraries of parts for fine-tuning gene circuit expression (24–28). As such, *B. subtilis* is an ideal candidate for the development of a well-characterized minimal chassis in which complex metabolic and regulatory circuits can be implemented and fine-tuned. First, we applied a greedy genome reduction strategy in which deletions in large dispensable regions were serially combined starting from the largest. At each step, the fittest clonal population was used for the next deletion cycle. This strategy generated a synthetic evolutionary tree of 298 strains with reduced genomes, the smallest genome (of 2.73 Mb) having lost 1.48 Mb in size. Physiological phenotyping of the complete synthetic evolutionary tree showed that growth was progressively impaired with genome reduction as synthetic-sick interactions accumulated. In genome-reduced strains, most differentially expressed genes could be explained by the known transcription regulatory structure in *B. subtilis.* These strains also exhibited transcriptional signatures of a constitutive thiol-induced oxidative stress. In sharp contrast to genome-reduced *E. coli* (15), genome-reduced *B. subtilis* strains exhibited a decreased frequency of spontaneous mutation (up to 100-fold) and, with such a reduced mutation rate, short-term adaptive evolution by serial transfer did not readily allow growth recovery. However, inactivation of the mismatch repair system increased the mutation rate and allowed for the selection of competitive genome-reduced isolates, hence providing an engineering framework for efficient genome streamlining of chassis strains for synthetic biology applications.

MATERIALS AND METHODS

Media composition

Bacillus subtilis strains were grown at 37°C in lysogeny broth (LB) and in NMS, a rich medium of chemically defined composition. NMS is based on the minimal salts medium (10.8 g.l^{-1} K_2HPO_4, 6 g.l^{-1} of KH_2PO_4, 1 g.l^{-1} of Na Citrate $2H_2O$, 2 g.l^{-1} of K_2SO_4) supplemented with 0.4% glucose, 0.1% casamino acids (Difco Casamino Acids, Bacto), 0.01% L-tryptophane, 0.016% L-glutamine, 0.005% L-asparagine, 0.004% L-cysteine, 0.01% L-histidine, trace elements (0.001 g.l^{-1} of $MnCl_2{\cdot}4H_2O$, 0.0017 g.l^{-1} of $ZnCl_2$, 0.00043 g.l^{-1} of $CuCl_2{\cdot}2H_2O$, 0.0006 g.l^{-1} of $CoCl_2{\cdot}\ 6H_2O$ and 0.0006 g.l^{-1} of $Na_2MoO_4{\cdot}2H_2O$), 0.1 mM of $FeCl_3$, 0.1 mM of $CaCl_2$, 1 mM of $MgSO_4$ and 1 mg.l^{-1} of each following vitamins: B12, calcium pantothenate, nicotinic acid, pyridoxal, thiamine, folic acid, biotin and riboflavin. Solid media were obtained by adding 1.5% agar to the NMS liquid media. *B. subtilis* competent cells were transformed as described in (29). The antibiotics phleomycin, neomycin, chloramphenicol, erythromycin and spectinomycin were added to NMS or LB medium at final concentrations of 8, 15, 5, 0.5 and 100 μg.ml^{-1}, respectively. M9 minimal salts medium (8.5 g.l^{-1} Na_2HPO_4, 3 g.l^{-1} of KH_2PO_4, 1 g.l^{-1} of ammonium chloride, 0.5 g.l^{-1} of NaCl) supplemented with 0.5% glucose, 0.005% L-tryptophane, 0.1 mM $CaCl_2$, 1 mM of $MgSO_4$ and trace elements was used for viability tests. M9 minimal salts medium supplemented with 0.3% glucose, 0.005% L-tryptophane, 0.1 mM $CaCl_2$, 0.05 mM $FeCl_3$, 1 mM of $MgSO_4$ and trace elements (M9B medium) was used for transcriptome analyses. DSM (Difco Sporulation Medium) was supplemented with $MnCl_2$ (10 nM), $Ca(NO_3)_2$ (1 mM), $FeSO_4$ (1 nM) for sporulation assays.

Making deletions in the *B. subtilis* chromosome

Deletions were made in the master strain (MS), which is derived from the TF8A strain (30) lacking 233.4 kb of the chromosome relative to wild type *B. subtilis* 168 as previously defined in (10). Deleted regions include the prophages SPβ and PBSX, the prophage-like element *skin* (30), and the mobile element ICE*Bs1* (31), which excised spontaneously in our TF8A lab stock. The *upp* gene of TF8A was replaced with a neomycin-resistance gene under the control of the Lambda pR promoter (λpR-*neo*) to give the master strain TF8A λpR-*neo::Δupp* (10). All deletions were introduced in the master strain by homologous replacement of the targeted chromosome region by a DNA fragment called the [*upp ble λcI*] deletion cassette (Supplementary Figure S1), carrying the phleomycin-resistance gene for positive selection of cassette integration, and both the *upp* and P_{sak} *λcI* genes for counter-selection and cassette eviction. The primer pairs P1–P2 and P3–P4 (0.6 μM final each) (Supplementary Table S1) are mixed with ∼200 ng of master strain chromosomal DNA and amplified by PCR under standard conditions to generate DNA fragments at least 1.4 kb long (Supplementary Figure S1). The PCR products are treated with exonuclease I and shrimp alkaline phosphatase (Amersham) at 37°C for 60 min to digest the excess of primers and then the enzymes are inactivated by heating at 94°C for 10 min. The [*upp ble λcI*] cassette is amplified using Phleo3 and Phleo5 primer pair. Equal amounts (200 ng) of the DNA fragments P1–P2, P3–P4 and [*upp ble λcI*] cassette are mixed together and subjected to the joining PCR reaction under the following conditions: 5 min at 94°C; (10 s at 94°C, 10 s at 55°C, 12 min at 65°C) for 12 cycles; (10 s at 94°C, 10 s at 55°C, 12 min + 15 s cycle^{-1} at 65°C) for 24 cycles;

10 min at 72°C)] in the presence of P1 and P4 (0.2 μM final). Then, the joined products are dialyzed against water and used to transform competent cells of the master strain for phleomycin resistance on LB medium. Plates are incubated at 37°C up to 24 h, and eight independent colonies, preferably large and with a normal colony shape, are purified by streaking on the same selective medium, and the chromosome structure of an isolated colony was checked. The presence of the [*upp ble λcI*] cassette at the locus and the deletion of the chromosome interval are checked by PCR on the eight individual colonies using the primer pairs Phleo3-cI2RV and P5–P6, respectively (Supplementary Figure S1).

Deletion accumulation

Deletion accumulation in the lineage started with the combination of interval Δ1 with Δ4, Δ6 and Δ7. The three deletions (Δ4, Δ6 or Δ7) were respectively constructed with chloramphenicol (*cat*)-, erythromycin (*erm*)- and spectinomycin (*spc*)-resistant markers, which in contrast to the [*upp ble λcI*] cassette cannot be evicted by pop-out by lack of direct repeats (Supplementary Figure S1). The three deletions were successfully combined (i.e. by obtaining competent strain able to grow on NMS) as indicated on Supplementary Table S2. In such strain, one marker was then replaced by the [*upp ble λcI*] cassette, which was then evicted, and this operation was repeated to eliminate the two remaining selection markers. The same strategy was used for the construction of MGP022, MGP027, MGP029, MGP033, MGP046 and MGP052. Two versions of MGP052$^{p/o}$ (where $^{p/o}$ stands for pop out) fitted with either the *neo* (as for MS) or *cat* counter-selection marker were used to pursue genome reduction in two separate lineages. Then, deletion mutant strains were iteratively constructed with the [*upp ble λcI*] cassette.

Eviction of marker (pop-out) in deletion mutants

Deletion mutant strains with the expected chromosome structure are tested for cassette eviction. When the [*upp ble λcI*] cassette is present, the strain becomes sensitive to neomycin (Neo^S) because the λCI repressor inhibits the Lambda pR promoter which drives the transcription of the *neo* gene. The eviction of the [*upp ble λcI*] cassette from the chromosome occurs by a spontaneous single crossover recombination between 30 bp directly-repeated (DR) sequences which were included in the design of primer P3 (Supplementary Figure S1) (10,32). Such pop-out events are rare in the population, and they cause the cells to become resistant to neomycin and sensitive to phleomycin. Typically, cells are grown in LB without selection at 37°C until OD >2 and aliquots (100–1000 μl) are spread on LB plates supplemented with neomycin and incubated overnight at 37°C. A few Neo^R Phleo^S colonies are streaked on neomycin-containing plates. A well isolated Neo^R colony from each streak is used to verify the eviction of the cassette by PCR using P1–P4 primer pair. One strain in which eviction of the cassette by pop-out is efficient is used for the next deletion step.

High-throughput phenotyping of genome-reduced strains

Growth was determined by measuring light scattering of microtiter cultures using the BioLector (m2p labs, Baesweiler, Germany). This device controlled the temperature and humidity of the chamber and was able to measure light scattering without a pause in shaking, enhancing aeration. Clear-bottom microtiter plates with a well volume of 300 μl and a culture volume of 150 μl were used. Frozen glycerol stocks were thawed and plated on LB medium with appropriate antibiotic for one day, and 1 ml NMS precultures in 96-well deep well (2 ml) plates were inoculated from these colonies. After 24 h these were diluted 1/20 into fresh prewarmed NMS and immediately placed in the BioLector. All growth was done at 37°C. Because the initial phase typically involved a decline in light scatter for unknown reasons, light scattering data were normalized to the lowest recorded value for each well. This normalization produced a good fit to an exponential growth curve for the growth phase. The conversion between light scattering and standard $OD_{600\ nm}$ values was determined by using a dilution series of concentrated cells and found to be approximately linear.

RNA preparation and transcriptome analysis

A single colony of *B. subtilis* strains, MS, MGP192, MGP229, MGP254, MGP184 was grown in 10 ml of LB medium for 2 h. The cultures were then diluted 8000-fold in NMS medium and were incubated with shaking at 37°C until they reached an $OD_{600\ nm}$ of 0.3. Total RNA was isolated by acid-phenol extraction as described in (28). For transcriptome analysis, 35 μg RNA were DNase-treated using the RNase-Free DNase Set (Qiagen) and purified using the RNA Clean-Up and Concentration Micro Kit (Norgen). The RNA concentration was measured using a NanoDrop ND-1000 (Thermo scientific, USA) spectrophotometer; the quality of the RNA preparations was assessed by means of the Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA) according to the manufacturer's instructions. Labelling of the sample and hybridization were performed in strand specific conditions by NimbleGen, as previously described (28). Data processing is described in Supplementary Materials and Methods. Raw data and gene-level expression values after between-sample normalization have been deposited in GEO database (GSE207089). Detailed results of the differential expression analysis are provided in Supplementary Table S3. High-resolution expression profiles along the genome can be explored interactively using the Genoscapist data browser (33).

Short-term evolution of MS, MGP192, MGP229, MGP254 and MGP229$^{p/o}$ Δ*mutS* strains

The *mutS* gene was deleted in MS and MGP229$^{p/o}$. The gDNA of the strain BKK17040 (*trpC2* Δ*mutS*::*kan*) (23) was used to transform MS, MGP229$^{p/o}$ and the transformants were selected on kanamycin. The constructed strains were sequenced. The strains were grown in NMS or LB at 37°C for >200 generations. Cultures were performed in 5 ml in tubes or 100 μl in microplates and diluted 1:30 or 1:50 into fresh medium two to three times a day. MGP229$^{p/o}$ Δ*mutS* was grown in LB for about 120 generations, and

then the cell population was diluted into two batch cultures which were evolved independently for 80 additional generations. Colonies from the short-term evolution were isolated and characterized (genome sequencing and fitness determination, Supplementary Material and Methods).

Mitomycin C sensitivity assay

Mitomycin C sensitivity was determined on bacterial cell cultures grown in LB. Serial dilution of bacterial cell cultures were spotted on LB plates and on LB plates supplemented with mitomycin C 60 ng.ml^{-1}, 80 ng.ml^{-1}, 100 ng.ml^{-1}. Mitomycin resistant colonies were counted after 48 h of incubation at 37°C. The rate was estimated by dividing the number of resistant CFU by the total number of CFU on LB.

Rifampicin resistant frequency

Overnight bacterial cultures were diluted 100-fold in fresh LB medium. Rifampicin resistant frequency was assessed on bacterial cell cultures grown in LB at 37°C until late exponential phase (∼6 generations). The cultures were plated on LB plates and on LB plates supplemented with rifampicin 10 μg.ml^{-1} and the colonies counted after 48 h of incubation at 37°C. The frequency was estimated by dividing the number of resistant CFU by the total number of CFU on LB.

RESULTS

Greedy reduction of the *B. subtilis* genome

We have previously built a library of 146 genome-reduced strains, each carrying the single deletion of a dispensable chromosome region, and together covering ∼76% of the *B. subtilis* 168 genome (10). The forty-eight largest dispensable regions (from ∼131 kb down to ∼16 kb), which cover 57 % of the genome, were selected for stepwise deletion from the chromosome (Figure 1A). These regions are ranked by size and hereafter named by rank number (from 1 to 48) (Table 1). Eight subregions resulting from the splitting of the second and third largest dispensable regions were also included and named by the rank number of the parent region associated with a letter (Table 1). Each of the 56 large dispensable regions was deleted from *B. subtilis* 168-derivative Master Strain (MS) genome (Supplementary Table S2). Two of the resulting strains harboring deletion of rank 5 (Δ5) or 18 (Δ18) exhibited severe growth defects on rich, chemically-defined NMS medium and were thus excluded from the genome-reduction combinatorial strategy. Next, deletions were accumulated by iteration of a two-step deletion cycle (Figure 1B), which consisted in: (i) deletion of a chromosomal region by homologous replacement with a selectable cassette (i.e. [*upp ble λcI*]); (ii) eviction of the cassette by spontaneous pop-out recombination to generate a marker-free deletion, hence producing a genome-reduced strain ready for another deletion cycle (Supplementary Figure S1). Beyond the substantial loss of functional protein-coding and -noncoding genes, deletions may lead to transcriptional deregulation in *cis* by remolding transcription units (TUs). Indeed, although deletion endpoints were designed to limit collision between TUs (10), a deletion may fuse previously independent TUs and lead to deregulation of the downstream TU through putative overexpression (for co-directional TU) or silencing (for TU in opposite direction) (Supplementary Table S4). Some deletions removed primary sequences involved in replication termination, potentially changing where replication forks disassemble (Supplementary Results 2.1), while others resulted in the emergence of new open reading frames (Supplementary Results 2.3 and Supplementary Table S4). To mitigate the risk of creating synthetic lethality, slow growth and loss of competence, several combinations of deletions were attempted in parallel at each cycle and a deletion mutant exhibiting good colony growth and competence was chosen for the next deletion cycle. Hence, by retaining a fit deletion mutant at each cycle, our strategy, potentially selects for the accumulation of spontaneous mutations that limits fitness loss.

A synthetic evolutionary tree of 298 strains with genomes reduced by up to 1.48 mb

Greedy deletion accumulation was initiated by the deletion of the largest interval (Δ1). Three double deletion mutants, Δ1,3, Δ1,4 and Δ1,6, were obtained after the second deletion cycle but not Δ1,2 and Δ1,5 (Figure 1C). In order to detect the most suitable combinations as early as possible in the deletion cycle, we accumulated up to three deletions marked with different antibiotic resistance genes. Three deletions were successfully combined in Δ1 (i.e. by obtaining competent strain able to grow on NMS) as indicated on Supplementary Table S2. In such strain (for instance strain Δ1,4,7,6; Figure 1C), selection markers were iteratively evicted. The strain containing the marker-free deletions was used to start a new cycle of accumulation of marked deletions. As this look-ahead approach became labor-intensive due to lower frequency of marker eviction, we decided after the introduction of the 11th deletion (Δ19, MGP052, ∼3.3 Mb) to accumulate large deletions one-by-one. From MGP052$^{p/o}$ (where $^{p/o}$ stands for pop out) two separate lineages for genome reduction were pursued in which 10 and 17 deletions, respectively, were further accumulated by repeating the deletion cycle (Figure 1C, Supplementary Figure S2). Finally, we obtained MGP184 (∼2.87 Mb) and MGP254 (∼2.73 Mb), two strains respectively reduced by ∼1.34 and ∼1.48 Mb relative to the *B. subtilis* 168 strain (Figure 1D). The whole tree is shown in Supplementary Figure S2 and the corresponding genotypes are found in Supplementary Table S5.

The chromosome structure at the *locus* deleted last was verified by PCR before and after cassette eviction. Loci located nearby and deleted in previous cycles were verified also by PCR. Global chromosome structure was assessed during the construction of the synthetic evolutionary tree for several intermediate strains, including MS, MGP029, MGP052, MGP075, MGP116 and MGP229 using genomic DNA hybridization to tiled oligonucleotide arrays (Figure 1C). These experiments confirmed that targeted regions were deleted, revealed the spontaneous excision of ICE*Bs1* in MS, and suggested that the junctions

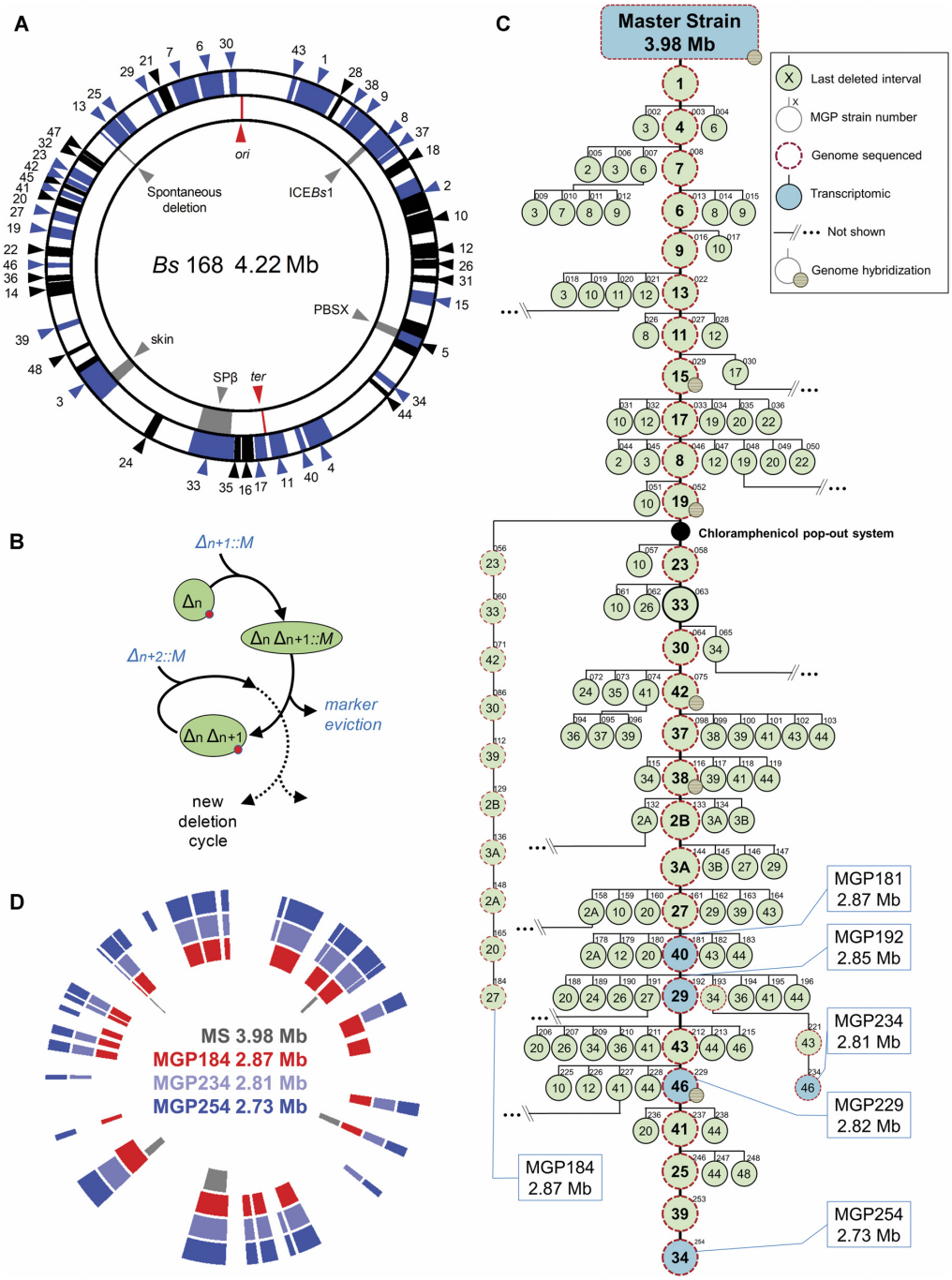

Figure 1. Iterative accumulation of large interval deletions in the *B. subtilis* chromosome. (**A**) The 48 largest intervals (mapped in black and blue) selected for stepwise deletion from the chromosome ranged from ∼131 to ∼16 kb and were named by rank number (Table 1 and Supplementary Table S4). The iterative deletion of the blue intervals led to MGP254. The ICE*Bs*1, PBSX, SPβ and skin elements (mapped in grey) have been previously deleted in the 168 *B. subtilis* background (30). The lab collection MS showed an additional spontaneous deletion (mapped in grey). *ori* and *ter* (mapped in red) correspond to the origin and terminus of replication, respectively. (**B**) The deletion cycle. A deletion mutant strain (Δ*n*) is transformed with DNA carrying a deletion of rank *n* + 1 linked to a selectable marker M (in blue). Upon its integration by homologous recombination, this DNA will transfer the deletion into the chromosome (see details in Supplementary Figure S1). Transformants showing the expected chromosome structure (i.e. deletion) at the targeted locus are subjected to marker eviction. Marker eviction occurs spontaneously by pop-out recombination between the short direct repeats and is selected by the concomitant eviction of λ*cI* and restauration of an antibiotic resistance. At this step, only strains that grow well on LB and NMS (green oval) are selected for another cycle of deletion. All the deleted loci were systematically verified by PCR at each cycle and after marker eviction. (**C**) Simplified lineage of deletion mutant strains (green circles, thick lines) showing the sequence in which deletions (indicated by rank number) have been introduced into the master strain to generate the strain with the largest genome reduction. The MGP254 minimal strain is poorly competent, preventing further deletion accumulation. Some of the deletion combinations attempted in parallel are shown by circles (thin lines). All the deletion combinations attempted in this study are shown in Supplementary Figure S2 and genotypes are given in Supplementary Table S5. Genome size of strains MGP181, MGP184, MGP192, MGP229, MGP234 and MGP254 are indicated on the side of the lineage. (**D**) Mapping of the individual deletions in MS, MGP184, MGP254 and MGP234 on a circular representation of the *B. subtilis* 168 chromosome (NC_000964.3). Each bar indicates a deleted genomic region.

Table 1. Dispensable regions used for genome reduction

Rank[a]	Size in MS (bp)	#CDS in MS	Interval genes end		168 coordinates[b]		Size in 168 (bp)	#CDS in 168
			First	Last	Start	End		
1	130 682	129	*ndhF*	*ycgG*	205 409	336 090	130 682	129
2	117 436	111	*yefA*	*yfkO*	737 603	855 038	117 436	111
3	99 202	108	*yqeF*	*greA*	2 647 920	2 792 217	144 298	169
4	87 147	32	*ymcC*	*ymzA*	1 781 228	1 868 374	87 147	32
5	83 485	87	*yjcM*	*ohrA*	1 263 702	1 381 403	117 702	134
6	83 883	87	*yxeH*	*rocR*	4 063 684	4 147 566	83 883	87
7	82 221	83	*yxlH*	*hutM*	3 966 745	4 048 965	82 221	83
8	58 318	68	*ydcL*	*ydgF*	529 417	608 245	78 829	93
9	77 116	79	*yclM*	*ydbO*	430 623	507 738	77 116	79
10	74 171	75	*yfkL*	*yfhO*	860 264	934 434	74 171	75
11	57 043	23	*yngA*	*yoeB*	1 946 140	2 003 182	57 043	23
12	56 115	54	*yhbF*	*yhdJ*	972 088	1 028 202	56 115	54
13	53 517	52	*lytD*	*ywqA*	3 684 826	3 738 342	53 517	52
14	49 886	48	*ytpR*	*ytwF*	3 052 743	3 102 628	49 886	48
15	46 370	54	*yisC*	*yitZ*	1 148 458	1 194 827	46 370	54
16	42 032	42	*yobI*	*yocS*	2 065 424	2 107 455	42 032	42
17	41 048	45	*yoxD*	*yobF*	2 018 591	2 059 638	41 048	45
18	40 107	40	*ydiQ*	*yebG*	658 183	698 289	40 107	40
19	38 495	32	*yueI*	*yuiA*	3 261 539	3 300 033	38 495	32
20	31 172	35	*yusO*	*yvrD*	3 374 492	3 405 663	31 172	35
21	36 453	37	*qoxD*	*ywaC*	3 914 273	3 950 725	36 453	37
22	35 307	33	*yubB*	*patB*	3 194 635	3 229 941	35 307	33
23	34 927	32	*araE*	*yvfA*	3 484 072	3 518 998	34 927	32
24	34 469	46	*ribT*	*yqkC*	2 427 392	2 461 860	34 469	46
25	33 405	40	*rapD*	*spoIID*	3 744 349	3 777 753	33 405	40
26	32 745	35	*yhdP*	*yhaP*	1 031 995	1 064 739	32 745	35
27	32 472	33	*yunB*	*yurT*	3 322 463	3 354 934	32 472	33
28	17 261	19	*ycgQ*	*yckE*	349 996	367 256	17 261	19
29	23 329	22	*ywfA*	*ywdH*	3 874 332	3 897 660	23 329	22
30	27 597	35	*yybP*	*yyaH*	4 169 134	4 196 730	27 597	35
31	26 801	25	*yhgD*	*yhxC*	1 089 755	1 116 555	26 801	25
32	26 379	25	*yvdM*	*yvcI*	3 546 828	3 573 206	26 379	25
33	25 004	34	*yodL*	*yplQ*	2 136 538	2 295 943	159 406	226
34	23 781	24	*yknT*	*abh*	1 494 363	1 518 143	23 781	24
35	23 633	26	*yojO*	*yodJ*	2 111 837	2 135 469	23 633	26
36	22 355	22	*ytvB*	*asnB*	3 105 470	3 127 824	22 355	22
37	21 732	24	*ydgI*	*ydhT*	612 191	633 922	21 732	24
38	21 672	23	*ycxA*	*yclH*	403 217	424 888	21 672	23
39	21 056	19	*ysgA*	*ysdB*	2 930 827	2 951 882	21 056	19
40	20 428	21	*ynxB*	*yncF*	1 880 087	1 900 514	20 428	21
41	18 974	19	*yvrL*	*yvgP*	3 411 624	3 430 597	18 974	19
42	18 435	21	*yvaQ*	*yvbK*	3 458 076	3 476 510	18 435	21
43	17 720	16	*ybaR*	*ybbK*	176 306	194 025	17 720	16
44	17 511	21	*slp*	*ylaK*	1 533 283	1 550 793	17 511	21
45	17 390	17	*yvgW*	*yvaM*	3 438 853	3 456 242	17 390	17
46	17 093	15	*ytdA*	*glgB*	3 154 554	3 171 646	17 093	15
47	16 906	16	*yvcE*	*yvpA*	3 574 363	3 591 268	16 906	16
48	15 752	17	*bofC*	*pheB*	2 836 909	2 852 660	15 752	17
2A	88 112	80	*yefA*	*yfmD*	737 603	825 714	88 112	80
2B	52 549	52	*yefA*	*yetL*	737 603	790 151	52 549	52
2C	64 762	59	*yetM*	*yfkO*	790 316	855 077	64 762	59
2D	86 932	87	*yesM*	*yfkO*	768 146	855 077	86 932	87
3A	74 942	85	*yqeF*	*aapA*	2 647 920	2 767 957	120 038	146
3B	49 211	58	*yqeF*	*yrpE*	2 647 920	2 742 226	94 307	119
3C	49 967	50	*sigZ*	*greA*	2 742 251	2 792 217	49 967	50
3D	75 247	78	*blt*	*greA*	2 716 971	2 792 217	75 247	78

[a] Δ5 and Δ18 were excluded from the combinatorial strategy as each independent deletion showed too important growth defects.
[b] according to the AL009126.3 genome sequence on GenBank.

of a few deletions were not located as expected. To identify potential unwanted rearrangements, the genomes of forty-three strains, including MS, and the two end-point strains MGP184 and MGP254, were fully sequenced. The MGP184 and MGP254 genomes harbored the expected 21 and 28 large deletions, respectively, and did not display any unwanted large-scale rearrangement, confirming that 1316 and 1506 annotated genes, respectively, were removed relative to the 168 reference genome (Figure 1D, Table 1). Among the 59 (MGP184) and 79 (MGP254) additional mutations identified, 55 and 75 mutations, respectively, were most likely of technical origin as they overlap regions covered by primers, and/or amplified by PCR, used for interval deletions. These mutations were indeed first detected in the genome sequence of the strains in which the corresponding interval had been removed by pop-in deletion or during cassette eviction by pop-out recombination. Of note, 28 of these mutations were common between MGP184 and MGP254, as they were inherited from the MGP052 common ancestor (Supplementary Table S6). The remaining mutations corresponded to scars generated by illegitimate recombination (Supplementary Figure S3), spontaneous deletion between naturally occurring direct repeats, and other mutations (point mutations and small indels). Overall, the two end-point strains MGP184 and MGP254 had slightly larger genomes than the previously published *B. subtilis* PG10 and PS38 genome-reduced strains (12). A genome alignment of MGP184, MGP254, PG10 and PS38 shows substantial overlap of deleted regions (Supplementary Figure S4A). It also highlights the distinct strategies used for genome reduction: the combination of many small deletions to minimize fitness loss in PG10 and PS38 (with an average deletion size of ∼17.2 and ∼16.3 kb, respectively) versus the systematic deletion of much larger regions for MGP184 and MGP254 (with an average deletion size of ∼64.0 and ∼53.0 kb, respectively). As a result, the genomes of our MGP184 and MGP254 strains have gene contents substantially distinct from that of previously reported genome-reduced strains (Supplementary Figure S4B, Supplementary Results 2.5).

Loss of function affects fitness of genome-reduced strains

We evaluated how the cell fitness varies upon genome reduction by determining growth rates and biomass titers in complex medium (Supplementary Figure S5). We analyzed the 272 strains carrying a single deletion of dispensable regions (10), and revealed that ∼24% and ∼40% of these strains exhibited a significant increase or reduction of growth rate relative to MS, respectively (Supplementary Table S5). Only ∼1% of the 272 single deletion strains exhibited a significantly increased biomass production, while 72% of these strains showed a significantly decreased biomass production indicating potential inability to consume the carbon source entirely or increased cell maintenance (Supplementary Figure S5C). Upon accumulation of deletions into the genome (Supplementary Figure S1C), virtually no strain with large deletions showed improvement over the wild-type by any fitness metrics (Figure 2A). Rather, cell fitness decreased as deletions accumulated into the genome. The decrease in cell fitness did not correlate with the change in *oriC-terC* symmetry, which ranged from +2° down to –17° as compared to MS (Supplementary Figure S5D, Supplementary Results 2.1).

In the whole library of strains that includes single and multiple deletion mutants, 42 large regions were by construction deleted at least three times independently. Among these, 10 deleted regions caused significant growth rate defects and lower biomass production relative to the parent strain (Δ2, Δ8, Δ12, Δ19, Δ20, Δ22, Δ24, Δ33, Δ36 and Δ45), while only two (Δ17, Δ23) showed a growth rate improvement correlated with higher biomass production relative to the parent strain (Figure 2B, Supplementary Figure S5F). Using knowledge from a large-scale study on single gene knockouts that affect growth on chemically defined media (23), we correctly predicted the positive, neutral, or negative effects on growth rate for 75% of the large deletions (Supplementary Table S4). For instance, the most significant growth defects were observed for the Δ20 and Δ22 strains (Figure 2B). Among inactivated genes, *citG* (in Δ20) and *pgi* (in Δ22) encode key enzymes of the central carbon metabolism and single deletions of these genes were shown to reduce growth significantly in a previous study (23). We found that growth defects in single gene deletion mutants predicted 72.7% of the growth phenotypes of 44 of our large deletion mutants (Supplementary Results 2.2), indicating that loss of function is likely the predominant cause of growth impairment in the genome-reduced strains. Our results also pointed significant growth defects that could not be predicted correctly. For example, the Δ2 strain exhibited an unpredicted growth defect but not the split Δ2A and Δ2B derivative strains, suggesting that at least 2 genes located in different parts of the Δ2 region are causing a synthetic-sick phenotype. Such synthetic-sick interactions are therefore expected to be revealed upon accumulation of deletions.

Synthetic–sick interactions accumulate upon genome reduction within a lineage

To decipher the consequences of the accumulation of synthetic interactions in a strain and to assess the extent to which these interactions are cumulative, we analyzed accumulation of interval deletions within the two main lineages (MGP184 and MGP254). Consistent with the previous analysis, Δ8 in MGP254 and MGP184 lineages and Δ20 in the MGP184 lineage resulted in severe growth impairments, while Δ17 and Δ23 resulted in better fitness (Figure 2C and D). In contrast, some deletions did not affect growth and fitness by themselves (Figure 2B) but resulted in severe growth defects when combined with other deletions. Examples include Δ9 and Δ15 in the MGP184 and MGP254 lineages (Figure 2C and D) and Δ2B in the MGP254 lineage (Figure 2C). Within each lineage, the accumulation of deletions globally resulted in cumulated synthetic-sick interactions (yellow zones in Figure 2C and D), with perhaps an exception for Δ20 in the MGP184 lineage for which the growth defect was lower than expected based on the single Δ20 mutant (green zone in Figure 2D). Interestingly, when introduced in a lineage some deletions (Δ7, Δ13, Δ27 and Δ33) appeared to set the growth rate to the level of the single deletion mutant, potentially offset-

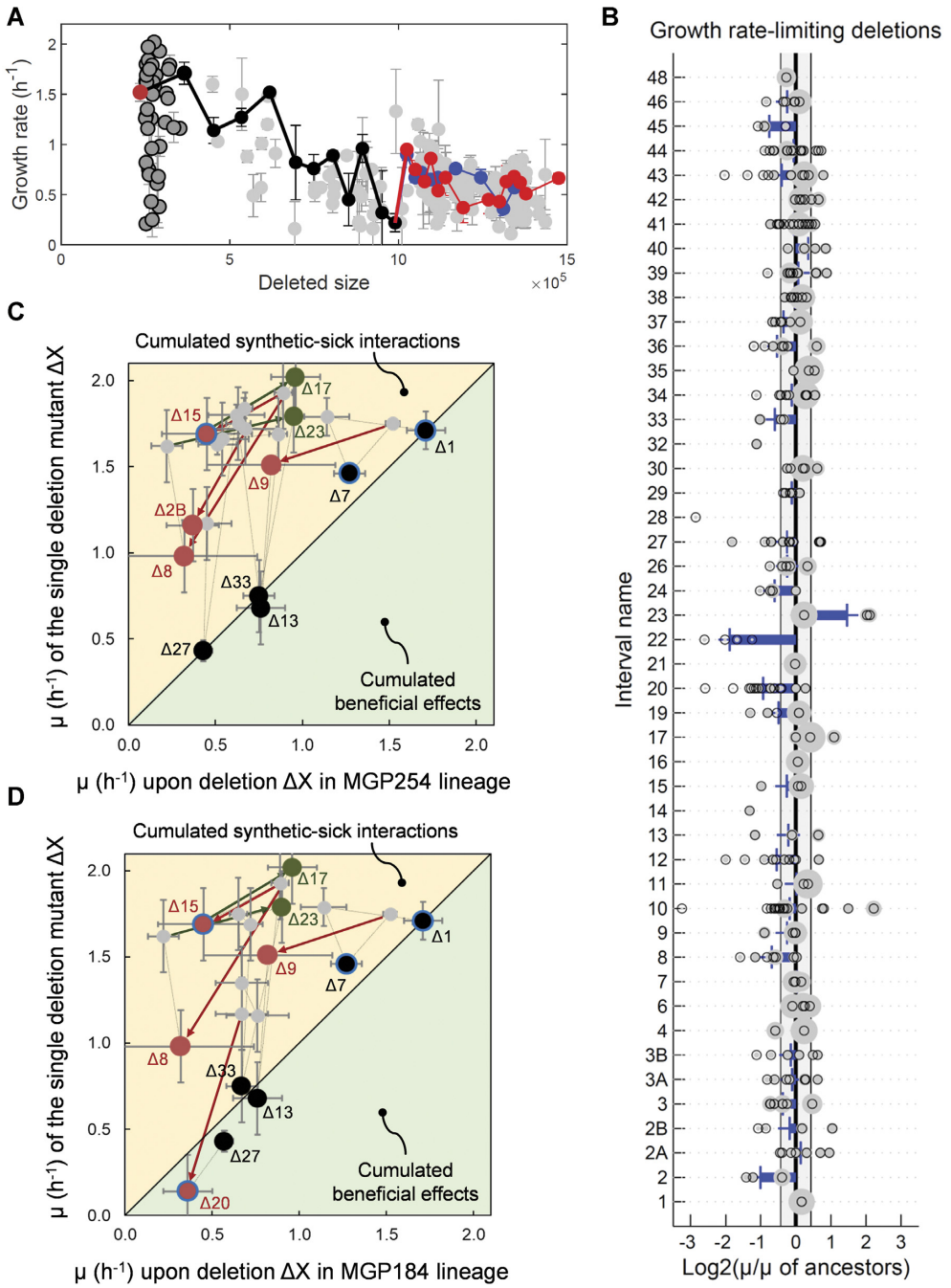

Figure 2. High-throughput phenotyping of the 298 genome-reduced strains. (**A**) Growth rate is negatively correlated to the deleted-genome size. Dots represent the single interval deletion (dark grey) and multiple interval deletion (light grey, black, blue and red) strains from MS (dark red). Edges represent the lineages: from MS to the most recent common ancestor (MGP052) of MG184 and MGP254 (black), from MGP052 to MGP184 (blue) and to MGP254 (red). The 95% confidence intervals are given as vertical bars. (**B**) Variation in $\log_2$ scale of the growth rate between a genome-reduced strain and its parental strain (circle) for the deletion of a given interval. Intervals were considered only if they were deleted in at least three strains within the 298 genome-reduced strain library. Diameter of the superimposed grey dots is proportional to the growth rate of the parental strain. For each interval is given the median variation of growth rate (horizontal blue bar). (**C**, **D**) Synthetic-sick interactions accumulate within MGP254 (C) and MGP184 (D) lineages as illustrated by plotting the growth rate of the strain deleted last for an interval (ΔX) in the two lineages *versus* the growth rate of the ΔX single interval mutant strains. Edges represents parental relationships. Green and yellow areas represent the cumulated beneficial effect and cumulated synthetic-sick interaction zones, respectively. Black dots on the diagonal correspond to interval deletions that cause a dominant phenotype. Green and red dots show interval deletions that led to significant growth rate improvement and reduction, respectively. Edges with arrowheads highlight the parental relationship between strains showing a significant reduction (red) or increase (green) of growth rate. Blue circled dots show intervals for which genome-scale, FBA-based analysis predicted a decrease in the growth rate upon accumulation in the MGP254 (C) and MGP184 (D) lineages, albeit much lower than experimentally measured. The 95% confidence intervals are given as vertical bars for the growth rates of single deletion mutants and as horizontal bars for the corresponding interval deletion within the lineages.

ting some of the previously cumulated synthetic-sick interactions (diagonals in Figure 2C and D). Altogether, these results reveal highly complex synthetic-sick interactions between large deletions.

We used our previously published metabolic model of *B. subtilis*, i*Bsu*1103 (34) to investigate whether synthetic-sick interactions could be explained by loss of known metabolic functions that accumulate through the deletion cycles. This model, further curated in (10), includes 1453 reactions associated with 1109 genes, and is one of the most complete models of *B. subtilis* available to date. Among the 1521 genes deleted in MGP254, 281 genes are present in i*Bsu*1103. These are associated to 366 reactions, and *in silico* deletion of each gene inactivated 201 reactions. During successive accumulation of deletions in the MGP254 lineage, flux balance analysis (FBA) of the model predicted a gradual reduction of fitness (associated with Δ1, Δ7 and Δ15; blue-circled dots in Figure 2C) until addition of Δ46, which is predicted to be lethal in chemically-defined medium. The predicted synthetic lethality arose from the loss of BSU30870 (Δ46) and BSU38860 (Δ7) resulting in UDP-glucose epimerase deficiency. Observed viability of the Δ1 Δ7 Δ15 Δ46 quadruple mutant in MGP254 lineage prompted us to revise the model. Indeed, the BSU07280 (*yfnG*) gene, which is not included in the model, could have a putative compensatory UDP-glucose epimerase activity, despite its low level of expression during exponential growth (28). This is supported by the fact that we were unable to construct the {(Δ2 OR Δ2A OR Δ2C OR Δ2D) AND (Δ7) AND (Δ46)} combination of deletions (Supplementary Figure S2), since Δ2, Δ2A, Δ2C and Δ2D deletions remove *yfnG* (Table 1). Using a model including the putative *yfnG*-encoded function, an additional reduction of fitness was predicted for Δ41. Overall, growth rates relative to MS were predicted with this revised model to decrease by 36.6% for MGP254 and 28.5% for MGP184 (Supplementary Figure S5E). These predicted endpoint values largely underestimated the experimentally observed reduction of about 65% of the growth rate for both MGP254 and MGP184 (Figure 2A). In addition, predicted fitness variations did not correlate to observed fitness variations. These discrepancies suggest that any potential fitness gain expected from the removal of dispensable cellular pathways (potentially alleviating cellular costs) is completely offset by the perturbation of cellular functions important for the overall cell fitness.

Genome reduction does not cause massive transcription deregulation

To investigate which cellular functions may be perturbed, five genome-reduced strains from the longest lineages (MGP181, MGP192, MGP229, MGP234 and MGP254) and the parental strain (MS) were subjected to transcriptome profiling during growth on NMS. We expected that impaired cell growth and cumulated deletion of genes exhibiting a wide range of expression levels in the parental MS strain would result in an extensive deregulation of transcription. However, with a cut-off on fold change (FC) of 1.5 ($|\log_2(FC)|>\log_2(1.5)$), ∼83% and ∼72% of the genes present in all genome-reduced strains (hereafter called the core genome) did not show any significant differential expression (DE) in the comparison of MGP181 and MGP254 to MS, respectively (Figure 3A). When considering DE genes with FC higher than 2 ($DE^{>2}$), core genes with unaltered expression increased up to 94% for MGP181 and 86% for MGP254 (Supplementary Table S3). The vast majority of genes from the core genome displayed no differential expression. Similar results were obtained for MGP192 and MGP229 (Supplementary Figure S6). These findings illustrate the robustness of the transcriptional network against massive genome reduction and suggest that synthetic-sick interactions may result from deregulation of a small number of genes.

Hierarchical clustering of the expression values of the genes from the core genome was performed to identify deregulated genes. It revealed that a relatively small number of clusters captures all the highly deregulated genes (Figure 3A). Among the 3766 annotated features of the core genome, 3314 are part of known regulons, including the housekeeping sigma factor σ^A regulon (35). The regulons of most of the sigma factors (i.e. $\sigma^{ABEFGHKMWX}$) and of the major transcriptional regulators (i.e. AbrB, CcpA, CcpC, CcpN, CggR, CodY, ComK, Fnr, Fur, PhoP, PurR, PyrR, ResD, Spo0A, Spx, TnrA, WalR, *etc.*) were not significantly affected by genome reduction. However, a small number of regulons showed significant $DE^{>2}$ in some of the genome-reduced strains. The regulons controlled by sigma factors σ^O, σ^V and σ^I, transcriptional activators ComA, CysL, GltC and NadR, transcriptional repressors AraR, CatR, CtsR, IolR, LexA, LutR, PerR, RocR, YetL and Zur, and by the nucleoid-associated protein YlxR, were upregulated. In contrast, the regulons controlled by σ^D, the repressor NrdR and the activator RemA were downregulated (Figure 3A). Comparative transcriptome analysis of the genome-reduced strains along the lineages also revealed that several genes were $DE^{>2}$, independently of their currently known regulons (Supplementary Table S3). For instance, expression of *comK* decreased with genome reduction until its extinction in MGP254, which is consistent with the gradual loss of competence of the genome-reduced strains. In addition, 46% (82 out of 177) and 33% (130 out of 391) of the $DE^{>2}$ genes in MGP181 and MGP254, respectively, are not known to be under the control of other factors than σ^A. The $DE^{>2}$ of these genes, which were expected to be constitutive, reflects our partial knowledge of the *B. subtilis* regulatory network.

Loss of function and aberrant transcription at deletion boundaries explain more than half deregulated genes

There are ∼230 genes annotated as '*regulation of gene expression*' in the *B. subtilis* 168 genome (35). MGP181 and MGP254 have lost 85 and 93 of these genes, respectively. Of these, about two-thirds are predicted to be regulators, but their exact functions are not characterized. We therefore focused our analysis on genes coding for previously characterized regulators. Most of these known regulators act on genes that are co-located on the genome. Thus, they tend to be co-deleted with their cognate regulons in our experiments. The remaining functionally characterized regulators act on distal genes and their deletions produced the expected deregulations in genome-reduced strains. For

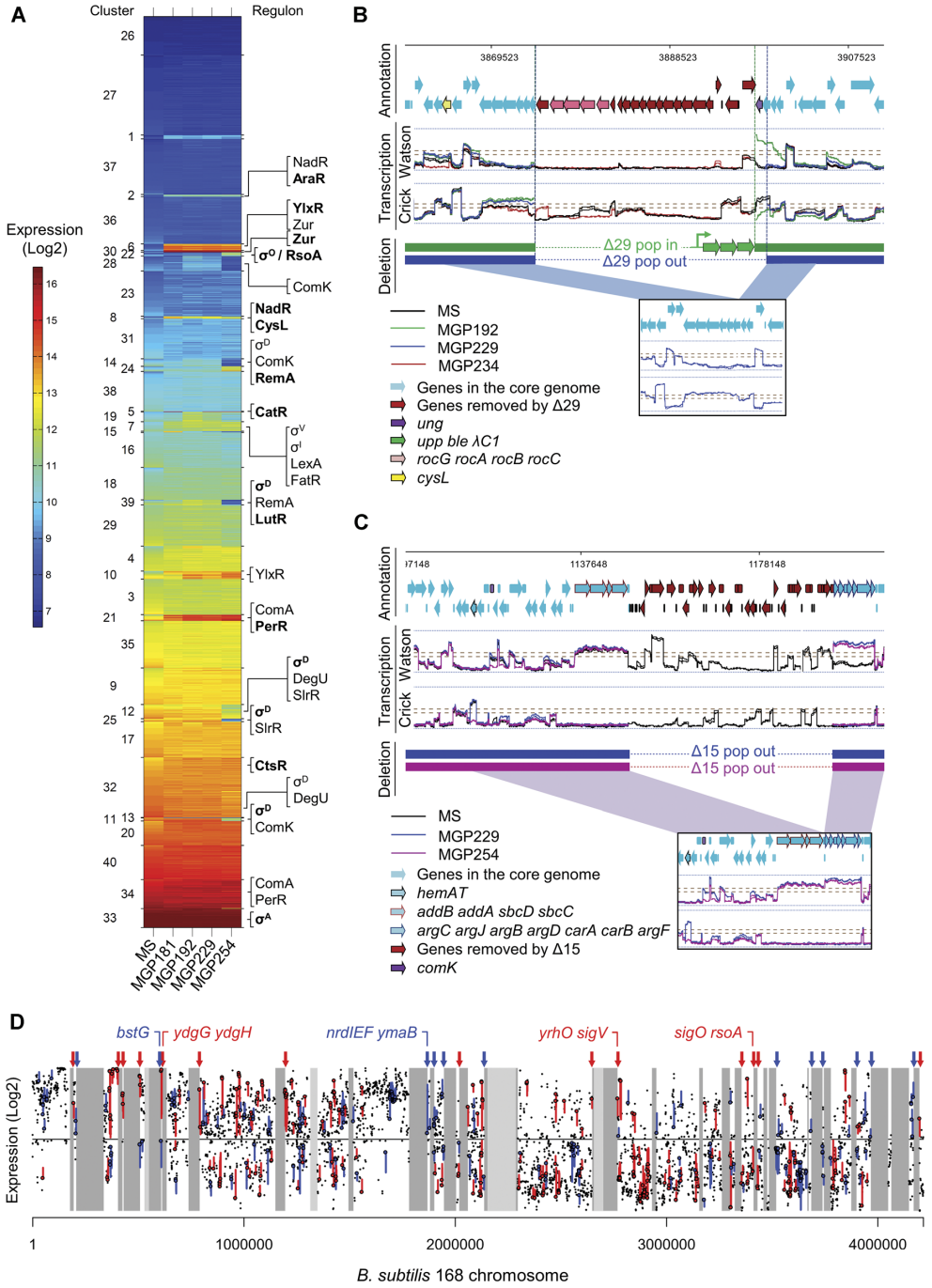

Figure 3. Transcription deregulation results from transcriptional conflicts and aberrant transcription at deletion boundaries and of loss of regulatory functions. (**A**) Hierarchical clustering (i.e. Ward linkage clustering based on Euclidean distance) of the expression level of the core genome of MS, MGP181, MGP192, MGP229 and MGP254 (in $\log_2$ scale). Cluster numbers are given on the left side of the figure (as indicated in Supplementary Table S3). Clusters enriched in regulons exhibiting highest differential expressions between strains are related to the σ^D, σ^O, σ^V, YlxR, NadR, Zur, PerR, ComA, *etc.* regulators (as indicated on the right side of the figure). Regulators in bold indicate a significantly high enrichment of the corresponding regulon within a cluster. (**B**, **C**) Genome coordinates, gene annotations and transcriptional landscape of genomic regions covering deletions Δ29 (B) and Δ15 (C), as viewed on the Genoscapist browser (33), specifically extended for representation of reduced genomes. The inset graph shows the transcriptional landscape after deletion of the corresponding interval, which resulted in the deletion of *ung* (B) and in synthetic transcripts (B and C). Genes and transcripts near deletion boundaries for which expression was strongly affected are indicated, such as the *cysL* antisense transcript S1458 (B) or the *comK* and *hemAT* transcripts (C). S1458 expression is similar in MS and MGP181, significantly increases in MGP192 and MGP229 but is not detectable anymore in MGP234 and MGP254. (**D**) Differential expression between MS and MGP254 mapped on the *B. subtilis* 168 chromosome. The deleted intervals are indicated as grey vertical rectangles (light grey for intervals absent in MS as compared to the 168 wild-type, and dark grey for intervals deleted in MGP254). Black dots indicate the level of expression of each gene (in $\log_2$ scale) in MS. Blue and red vertical segments show the downregulation and upregulation, respectively, of the corresponding genes between MS and MGP254. Vertical arrows on the top of the figure indicate downregulation (blue) and upregulation (red) attributable to transcriptional conflicts and aberrant transcription caused by the interval deletion. Examples of such deregulated genes are given as annotations of some vertical arrows.

example, Δ23 removes *araR* which encodes an arabinose-dependent transcriptional repressor, resulting in a strong upregulation of the AraR regulon (11 remaining genes in MGP181, MGP192, MGP229 and MGP234, but 1 in MGP254 as 10 were deleted in Δ39; Figure 3A). Similarly, the deletion of genes encoding the YetL (Δ2B) and IolR (Δ6) transcriptional repressors resulted in the upregulation of their respective regulons. Conversely, the combined loss of the RocR transcriptional activator (with Δ6) and the σ^L sigma factor (with Δ23) resulted in the expected strong downregulation of the σ^L-dependent RocR regulon (4 genes in MGP181 and MGP234; Figure 3A and B) (36). The deletion of *rocG* (Δ29; Figure 3B) further led to a GltC-dependent upregulation of *gltA* and *gltB* (37).

The removal of enzymatic functions also caused expected transcription deregulations. For instance, the deletion of the high-affinity zinc uptake system encoded by *znuA znuC znuB* (Δ1) and *zinT* (Δ3A) resulted in a substantial upregulation of the Zur regulon (7 genes still present in the core genome; Figure 3A). Deletion of the niacin transporter-encoding gene *niaP* (Δ1) resulted in the strong induction of the NadR regulon (5 genes still present in the core genome; Figure 3A; Supplementary Figure S7A). Deletion of the histidine permease HutM (Δ7) strongly upregulated 21 of the 23 genes (including the histidine and arginine biosynthesis cluster) that form the YlxR nucleoid-associated protein regulon, suggesting *de novo* synthesis of histidine by a yet unknown mechanism (Figure 3A).

The deletion of structural genes also led to transcription deregulations. For instance, several differentially regulated genes in MGP254 belonged to the '*exponential* and *early post-exponential lifestyle*' category, which groups genes involved in motility, biofilm formation, swarming, *etc*. Out of the 61 genes of this functional category, 40 σ^D-dependent and 10 RemA-dependent genes are downregulated in MGP254 (Figure 3A). The σ^D regulon is substantially affected in MGP254, with 91 out of 105 target genes strongly downregulated (such as *hemAT* as indicated on Figure 3C). The anti-sigma factor FlgM is secreted by the flagellar export apparatus in *B. subtilis* and its secretion relieves σ^D-dependent gene expression (38). In the MGP254 lineage, downstream MGP229, the structural flagellar genes *flhO flhP* were deleted (Δ25), and it has been previously shown that mutants defective in flagellar hook completion (and in particular the *flhO flhP* mutant) result in strongly reduced σ^D-dependent gene expression by lack of FlgM secretion (39). Altogether, one third of the genes differentially regulated in the genome-reduced strains can be explained by the loss of known gene functions (such as regulators, enzymes/transporters, and structural genes).

We also observed transcription deregulations at deletion boundaries. For instance, the insertion of the [*upp ble* λ*cI*] deletion cassette, which lacks a transcription terminator at its 3'-extremity, may lead to deregulation of downstream genes resulting from transcription read-through from the cassette (Supplementary Figure S1). Insertion of the cassette for deletion Δ29 in MGP192 created a synthetic non-coding transcript that likely acts as an antisense RNA for *ung*, resulting in its strong repression (Figure 3B). In addition, the eviction of the cassette led in this case to an illegitimate recombination and to the deletion of *ung* (Supplementary Figure S3; Supplementary Figure S7B). Artificial TUs were also created by the cassette eviction as illustrated by the fusion of the *addB* and *argC* operons following deletion Δ15 in MGP029 (Figure 3C). Consequently, the *argC* operon was constitutively expressed in the descendant strains. Large interval deletions also led to the deregulation of the expression of sigma factors at deletion boundaries. For instance, the deregulation of the RsoA/σ^O regulon results from either the deletion of *rsiO*, which encodes the σ^O anti-sigma, or the highly expressed *cysJ sigO* synthetic transcriptional fusion, both resulting from deletion Δ41 (Figure 3A and D; Supplementary Figure S7C). Together, we detected 26 transcriptional conflicts and aberrant transcription at deletion boundaries (i.e. *approx.* 20% of the $DE^{>2}$ genes) (Figure 3D; Supplementary Results 2.4).

Illegitimate recombination events and mutations, which are most likely of technical origin as they lie at deletion boundaries, also contributed to trans-transcriptional deregulation. For instance, illegitimate recombination upon the eviction of the cassette following deletion Δ9 resulted in the inactivation of *rapC* and deletions of *phrC*, *yczM* and *yczN* (Supplementary Figure S3). As transcription of the RapC- and RapF-encoding genes is positively regulated by their own target ComA, partly breaking the autoregulatory circuit led to the induction of the ComA regulon (10 out of the 15 genes present in the core genome, including *comS*, with only the targets of both ComA and FapR not induced; Figure 3A). A frameshift mutation in the *catR* gene at the boundary of Δ42 also resulted in the strong induction of the CatR regulon (*catDE*; Figure 3A). Altogether, our analysis of the transcriptomes of genome-reduced strains shows that over half of the observed transcriptional deregulations are attributable to the loss of metabolic and regulatory functions that trigger regulation cascades or are caused by side effects near deletion boundaries.

Genome-reduced strains exhibit thiol-induced oxidative stress

Transcriptome analyses revealed a striking common feature in our genome-reduced strains: the upregulation of genes involved in cell responses to oxidative stresses. Indeed, all genes from the core genome known to be under control of the transcriptional repressor PerR (13/13) were highly upregulated, including genes coding the catalase KatA, the heme biosynthesis enzymes, the iron storage protein MrgA, and the master regulator Spx (Figure 3A). Most (7/12) genes controlled by the CtsR repressor were also induced. CtsR regulates the production of the ATP-dependent protease complexes, ClpCP, ClpEP and ClpXP that act as protein quality control machineries, and of thioredoxin TrxA that mediates the reduction of proteins by cysteine thiol-disulfide exchange. The CtsR-dependent induction of *trxA*, together with the CtsR-independent induction of *trxB* encoding the NADPH-dependent thioredoxin reductase is characteristic of the thiol-oxidative stress response (Figure 4, Supplementary Table S3). The thioredoxin system acts as a redox buffer to maintain homeostasis and protect proteins during oxidative and thiol stresses (40), and is involved in Fe−S cluster metabolism as an effective reductant of persulfides formed as products of cysteine desul-

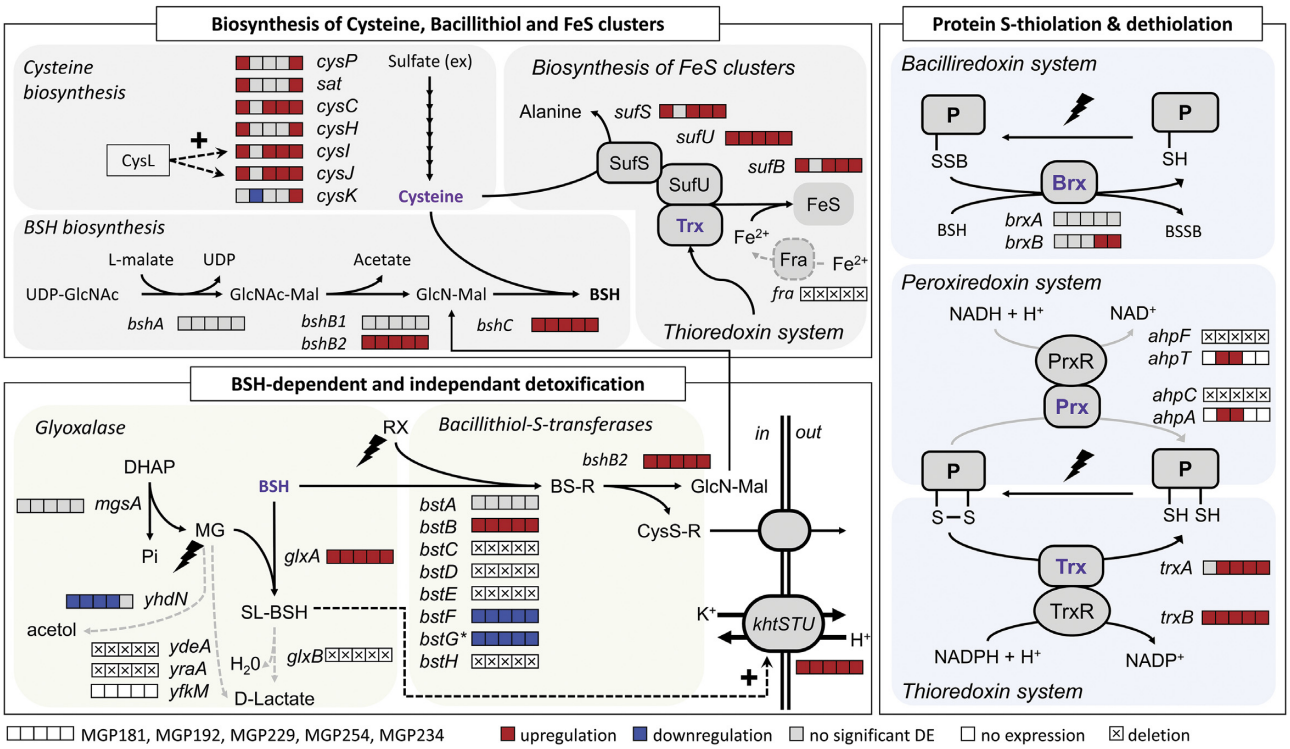

Figure 4. A genome reduction-induced oxidative thiol stress. Cysteine, bacillithiol (BSH), and the bacilliredoxin (Brx), thioredoxin (Trx) and peroxiredoxin (Prx) systems act as redox buffer to maintain homeostasis, protect proteins during oxidative and thiol stresses, detoxify the cell from small molecule compounds, and reduce protein persulfide inherent to the Fe−S cluster metabolism. Cysteine biosynthesis from extracellular sulfate involves seven reactions, two of which are under the control of CysL. The thioredoxin and BSH biosynthesis pathways are under the control of, among others, the major transcriptional regulator Spx (50,51). DE (as compared to MS) is given for each reaction-encoding gene by an array of boxes (from left to right: MGP181, MGP192, MGP229, MGP254, MGP234) colored in red for upregulation, blue for downregulation, grey for no significant DE, white for no expression, and ticked box when the gene has been deleted. The superscript * to *bstG* (also known as *dinB*), i.e. *bstG**, means that the observed gene expression deregulation is related to a transcriptional conflict at the Δ8 deletion boundary (Supplementary Figure S7D). UDP-GlcNAc, UDP-N-acetylglucosamine; GlcNAc-Mal, *N*-acetyl-α-D-glucosaminyl-L-malate; GlcN-Mal, D-glucosaminyl-L-malate; DHAP, dihydroxyacetone phosphate; Pi, inorganic phosphate; MG, methylglyoxalate; SL-BSH, S-lactoyl-BSH; RX, electrophiles, xenobiotics, and toxins; BS-R, BS-conjugates; CysS-R, cysteine conjugates; PrxR, peroxiredoxin reductase; TrxR, thioredoxin reductase.

furase reactions (Figure 4; (41)). Consistently, the *sufCDSUB* operon (biogenesis of Fe−S clusters) was upregulated in all genome-reduced strains. Concomitantly, the cysteine biosynthetic *cysJI* operon, whose transcription is controlled by the CysL activator, was significantly upregulated in all genome-reduced strains (except MGP192; Figure 4), potentially reflecting accumulation of sulfate, sulfite, or thiosulfate as described in conditions of high *cysJI* expression (42). In addition to cysteine and Trx, *B. subtilis* possesses a low-molecular-weight thiol system, bacillithiol (BSH, glycoside of L-cysteinyl-D-glucosamine with L-malic acid) to maintain homeostasis during oxidative and thiol stresses (43). Among BSH biosynthetic genes, *bshB2* and *bshC* were significantly upregulated in all genome-reduced strains (Supplementary Table S3), where *bshA* and *bshB1* were not $DE^{>2}$. Both *bshA* and *bshB1* are located in an operon with *mgsA* encoding a methylglyoxal (MG) synthase. MG is a toxic byproduct of glycolysis detoxified through both BSH-dependent (with glyoxalases GlxA and GlxB) and BSH-independent pathways (with YdeA, YraA, YfkM) that convert MG into D-lactate (Figure 4), and through the YhdN-dependent BSH-independent pathway that converts MG into acetol (44). The lack of *glxB* gene (Δ27) is known to provoke accumulation of the S-lactoyl-BSH intermediate and to activate the KhtSTU K + efflux pump to facilitate detoxification. As shown on Figure 4, the *khtSTU* genes were upregulated in all the genome-reduced strains whereas the BSH independent pathways for MG detoxification involving YfkM and YhdN were not expressed and downregulated, respectively. These observations are highly consistent with a general response to thiol-oxidative stress taking place in genome-reduced cells. A comparative analysis of DE genes in the genome-reduced *B. subtilis* strains constructed by Reuss *et al.* (12) revealed an upregulation of thiol-oxidative stress, in keeping with our observations, as well as a small set of deregulated genes common to all genome-reduced strains (Supplementary Table S7; Supplementary Results 2.6)

Genome-reduced strains are highly resistant to the DNA damaging mitomycin C

Our transcriptional analysis identified an upregulation of the LexA regulon (Supplementary Table S3). LexA controls the SOS response to DNA damage, and as a result leads to transient inhibition of cell division to allow DNA reparation, which in turn results in elongated cells (45). We therefore performed microscopic observations of the MGP229

and MGP254 genome-reduced strains. We observed elongated cells with altered cell morphology and perhaps defects in the spatial positioning of cell division sites and, most likely, the DNA replication and chromosome segregation machineries (Supplementary Figure S8, Supplementary Results 2.7). We next evaluated the sensitivity of MS, MGP192$^{p/o}$, MGP229$^{p/o}$ and MGP254$^{p/o}$ strains to the DNA damaging agent mitomycin C (MMC), which causes formation of double-stranded DNA breaks (DSBs) in the chromosome and induces the SOS response. Cells were deposited on plates in the presence or absence of MMC and survival ratio was assessed after incubation at 37°C (Figure 5A). Increasing amounts of MMC resulted in a drastic reduction of MS viability (up to 10^5-fold in the presence of 100 ng.ml^{-1} MMC). Strikingly, MGP192$^{p/o}$, MGP229$^{p/o}$ and MGP254$^{p/o}$ did not show any substantial loss of viability up to 80 ng.ml^{-1} MMC and a 10-fold reduction at 100 ng.ml^{-1} MMC. To determine how MMC sensitivity may be so drastically changed by genome reduction, 25 strains covering the lineage from MS to MGP254 were tested using the same assay (Figure 5B). From MS to MGP046 (last deletion, Δ8), resistance to MMC exhibited a near-logarithmic increase with the number of deleted intervals. The next strain in the lineage, MGP052 (last deletion, Δ19), significantly reverted (>10-fold) MMC resistance relative to MGP046. Of note, Δ19 creates a synthetic antisense RNA knocking-down expression of the *bstG* gene (formerly called *dinB*, for DNA damage-inducible (*din*) genes; Supplementary Figure S7D). The next deletion (Δ23) reverted MMC resistance a further 10-fold bringing MGP058 to the level of MS. From MGP058 to MGP075 (last deletion, Δ42), MMC resistance remained constant and similar to MS. Then, MMC resistance sharply increased (>300-fold) in strain MGP098 (last deletion, Δ37) and remained at this high level for all subsequent deletions in the lineage (from MGP116 to MGP254; Figure 5B). This profile of MMC resistance across the lineage did not correlate with cell growth rate, *ori-ter* symmetry disruption nor genome size (Supplementary Table S5). It highlights that complex beneficial- and synthetic-sick interactions that accumulate during genome reduction can lead to the emergence of striking new phenotypes. Importantly, because of these complex interactions, the mechanisms underlying MMC resistance in genome-reduced strains are currently unknown and do not necessarily reflect a greater capacity of cells to repair DSBs.

Genome-reduced strains exhibit extremely low spontaneous mutagenesis

Sequencing revealed only a few dozen mutations in the reduced genomes, of which >90% were of technical origin. As we were expecting an accumulation of spontaneous fitness-restoring mutations throughout the genome reduction process, we investigated spontaneous mutation frequencies in MS, MGP192$^{p/o}$, MGP229$^{p/o}$ and MGP254$^{p/o}$. Using a classical assay in which spontaneous mutations in the *rpoB* gene confer resistance to rifampicin, we determined the proportion of Rif^R mutants in the viable cell population. MS exhibited a frequency of Rif^R mutations of 4.1×10^{-8} (Table 2), which is similar to the mutation frequency in *B. subtilis* 168 (about 1×10^{-8}) previously reported under similar experimental conditions (46). Surprisingly, genome-reduced strains exhibited Rif^R mutation frequencies 10- to 100-fold lower than MS, with 4.1×10^{-9}, 4.4×10^{-10} and 2.8×10^{-9} for MGP254, MGP229 and MGP192, respectively. These results indicate that the genome-reduced strains exhibit significantly lower frequency of spontaneous mutation than the reference strain. The mismatch repair (MMR) system corrects errors that occur during DNA replication (47), and inactivation of MMR leads to a 100-fold increase in the mutation rate in *B. subtilis* (48). We deleted the *mutS* gene in MGP229$^{p/o}$ (due to the lack of competence of MGP254) and MS strains to inactivate the MMR system. We found that mutation frequencies in MGP229$^{p/o}$ Δ*mutS* and MS Δ*mutS* were similar and elevated by about 100-fold relative to MS ($2.5–3.5 \times 10^{-6}$; Table 2). These results may suggest that the low mutation frequency observed in MGP229$^{p/o}$ entirely depends on a functional MMR system. We then tested whether the emergent phenotypes of high resistance to MMC and of low spontaneous mutagenesis in MGP229$^{p/o}$ were related by assessing the sensitivity of the MGP229$^{p/o}$ Δ*mutS* strain to MMC (Figure 5C). We found that the MGP229$^{p/o}$ Δ*mutS* strain was as resistant to MMC as MGP229$^{p/o}$, indicating that the two emergent phenotypes observed in the genome-reduced strains are most likely independent. As *mutSL* genes have similar levels of expression in MS and MGP229$^{p/o}$ strains (Supplementary Table S3), our findings suggest that MMR may be more efficient in genome-reduced strains, drastically reducing spontaneous mutagenesis and probably limiting evolution.

Table 2. Mutation rates of the genome-reduced strains

Strain[a]	Mutation rate[b]
MS$_{cat}$	$4.1 \times 10^{-8} \pm 2.6 \times 10^{-8}$
MGP192$^{p/o}$	$2.8 \times 10^{-9} \pm 0.1 \times 10^{-9}$
MGP229$^{p/o}$	$4.4 \times 10^{-10} \pm 0.3 \times 10^{-10}$
MGP254	$4.1 \times 10^{-9} \pm 0.7 \times 10^{-9}$
MS$_{cat}$ Δ*mutS*	$2.4 \times 10^{-6} \pm 1.0 \times 10^{-6}$
MGP229$^{p/o}$ Δ*mutS*	$3.5 \times 10^{-6} \pm 1.0 \times 10^{-6}$

[a]MS$_{cat}$ is isogenic to MS, except that Δ*upp*::[λ*pR-neo*] has been replaced by Δ*upp*::[λ*pR-cat*]. The abbreviated genotype of MGP192$^{p/o}$ is MS$_{cat}$ Δ(1, 2B, 3A, 4, 6, 7, 8, 9, 11, 13, 15, 17, 19, 23, 27, 29, 30, 33, 37, 38, 40, 42)$^{p/o}$, that of MGP229$^{p/o}$ is MGP192$^{p/o}$ Δ(43, 46)$^{p/o}$ and that of MGP254 is MGP229$^{p/o}$ Δ(25, 39, 41)$^{p/o}$, Δ34::[*upp ble λcI*]. Complete genotypes are listed in Supplementary Table S2.
[b]Overnight cultures were spread (0.2 mL per plate) on LB supplemented with rifampicin (10 mg.ml^{-1}) and the number of viable cells was quantified on LB. The proportion of rifampicin cells per viable cells was calculated and the means and standard deviations are given from three independent experiments.

Genome-reduced strains exhibit a reduced capacity for short-term adaptive evolution

In order to test whether the cell fitness of genome-reduced strains can be enhanced by the accumulation of beneficial spontaneous mutations, we performed short-term adaptive laboratory evolution experiments using MS, MGP192, MGP229 and MGP254 as well as MGP229$^{p/o}$ Δ*mutS*, as the evolution of non-MMR-deficient genome-reduced

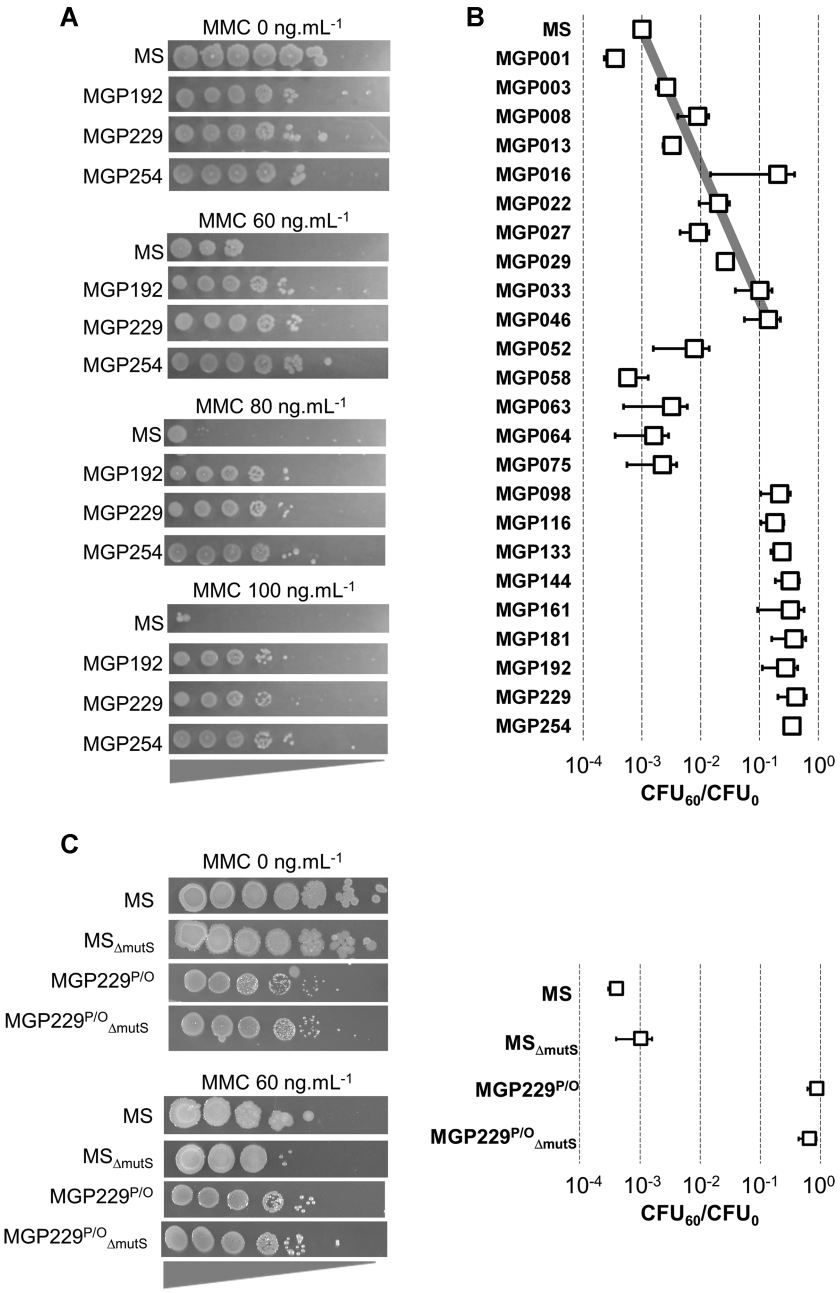

Figure 5. Genome-reduced strain sensitivity to mitomycin C. (**A**) Overnight cultures of MS, MGP192, MGP229 and MGP254 have been successively diluted 10-fold and dropped on LB plates supplemented with mitomycin at different concentrations (0, 60, 80 and 100 ng.ml^{-1}). After 48 h and 72 h incubation at 37°C plates were observed. The experiment was performed four times and the data show a representative result at 72 h. (**B**) Similar experiments were conducted on the lineage from MS to MGP254 in order to monitor mitomycin sensitivity along the lineage. The ratio (in log scale) of the number of CFU (Colonies Forming Unit) in the presence of 60 ng.ml^{-1} mitomycin by the number of CFU in the absence of mitomycin is given for each strain. Standard deviations of three independent experiments are shown as horizontal bars. The gray line is a logarithmic fit of MMC resistance versus the number of deleted intervals ($R^2 = 0.7$). (**C**) Sensitivities toward MMC of strains, MS, MS Δ*mutS*, MGP229$^{p/o}$, and MGP229$^{p/o}$ Δ*mutS* were evaluated as in panels A and B.

strains might be limited. Two independently evolved isolates MS^{evolA} and MS^{evolB} displayed increased growth rate and biomass production (by up to 34%, Table 3). Whole genome sequencing of the MS^{evolA} and MS^{evolB} isolates identified five and seven, respectively, point mutations, as well as small deletions (Supplementary Table S8). In contrast, none of the independent, short-term evolutions of MGP192, MGP229 and MGP254 yielded isolates with detectably increased growth rate or biomass production. We tested whether single mutations identified in MS^{evolA} and MS^{evolB} genome were able to improve growth rate and titer in naïve MS and genome-reduced MGP192$^{p/o}$ and MGP229$^{p/o}$ strains (Supplementary Results 2.8). Six mutant alleles in four genes were transferred in naïve MS. We found that the two alleles *nhaC*$_{\Delta A17}$ and *nhaC*$_{C515T}$ increased slightly but significantly growth rate and biomass

(by up to 10%; Table 3). The $sigB_{G451A}$ and $rsbW_{C70G}$ alleles did not increase MS growth rate but slightly improved biomass yield, which may be associated with the drop of sporulation efficiency caused by these alleles (Supplementary Table S9). These four alleles were then transferred in MGP192$^{p/o}$ and MGP229$^{p/o}$. We observed that none of these alleles improved growth rate detectably but that $nhaC_{\Delta A17}$ and $nhaC_{C515T}$ slightly increased biomass yield in MGP192$^{p/o}$ (Table 3). These results suggest that although single mutations may increase fitness in MGP192$^{p/o}$, the strain was not able to acquire such beneficial spontaneous mutations in our adaptive evolution experiments.

We also subjected to adaptive evolution the MGP229$^{p/o}$ Δ*mutS* strain in which spontaneous mutations occur at a high frequency (Table 2). Independent clones were evaluated for growth rate and biomass. A 36% increase in growth rate and a 17% increase in biomass were observed (Table 3), indicating that genome-reduced strains can produce more competitive variants under conditions where the mutation rate is increased by MMR inactivation. Two of these competitive variants (one from each partially independent lineage) were fully sequenced and 40 and 41 mutations were identified, with 28 mutations in common (Supplementary Table S8). Some of these mutations likely corrected previously identified transcriptional deregulations and synthetic-sick interactions. Indeed, the potentially deleterious consequences of overexpressed genes appeared to be corrected by inactivation of the gene product activity (e.g. a mutation near ArgH active site) or a frameshift mutation preventing translation (e.g. YvzA of unknown function). Also, the GltC-dependent transcriptional deregulation of the *gltAB* operon in the absence of *rocG* (Δ29; (37)) has been rectified by a frameshift mutation in *gltC*. These results prompted us to conclude that compensatory mutations in *B. subtilis* can at least partly correct the transcriptional and metabolic deregulations occurring upon genome reduction and allow for the selection of more competitive isolates. Hence, the apparent lack of evolution in genome-reduced strains is likely attributable to the yet unexplained low spontaneous mutagenesis.

DISCUSSION

Using custom-made genome engineering tools, we built a library of 298 genome-reduced *B. subtilis* strains by combining deletions of 48 large genomic intervals. The smallest genome, MGP254, lacks 1.48 Mb in size relative to the reference 168 strain (Figure 1). This strain has lost about 25% of the 1700 genes identified as evolutionary conserved at species-level by comparative genomic analyses of many *B. subtilis* genomes (49). A large fraction (72%) of the single interval deletions (272 strains) had detrimental effects on biomass production whereas only a few deletions (∼1%) slightly improved biomass production (Figure 2). Most of the detrimental effects (∼75%) could be explained by the loss of gene functions known to affect growth. Upon combination of these deletions, we observed synthetic-sick interactions as evidenced by synergistic detrimental effects between large deletions. Overall, genome reduction was associated with a progressive accumulation of synthetic-sick interactions, leading to a 65% decrease of cell fitness in MGP254 relative to MS. The i*Bsu*1103 metabolic model (34) refined using phenotypes from single large deletions (10) was unable to predict the observed fitness gains and losses throughout genome reduction, indicating that a large part of the interactions in the cell metabolic network remains to be uncovered.

Despite this accumulation of synthetic-sick interactions and strong effects on cell fitness caused by our greedy approach to genome reduction, we found that 86% of the genes present in all genome-reduced strains were not differentially expressed (<2-fold). To compare the transcriptomes of genome-reduced strains, we extended the Genoscapist data browser (33) to help visualize that only a relatively small number of regulons controlled by sigma factors, transcription activators and repressors were highly deregulated (Figure 3). Among observed deregulated genes, about one third correspond to regulons affected by the loss of their known transcription regulator and another 20% by aberrant transcription near deletion boundaries. These findings illustrate the robustness of the cell transcriptional network against massive genome reduction. They also helped identify a striking feature common to our genome-reduced strains, the upregulation of genes involved in response to oxidative stress, specifically cysteine, bacillithiol and Fe–S clusters biosynthesis, bacillithiol-dependent detoxification, and protein S-thiolation pathways (Figure 4). A significant upregulation of the CtsR regulon controlling the CtsR-dependent protein quality control machinery was also observed (Figure 3), indicating a cell response to protein misfolding. These upregulations cannot be explained by the currently known regulatory networks in a simple fashion. They are consistent with previous observations in *E. coli* and *B. subtilis* that genome reduction leads to global responses to oxidative stress and protein misfolding (12,14) (Supplementary Table S7). We do not know which deletion(s) upstream of MGP181 in the lineage may have caused the observed oxidative stress response. Possibly, such a stress response may arise from the cumulated loss of multiple cellular functions acting to control oxidative thiol stress. These functions comprise AhpC AhpF (alkyl hydroperoxide reductase subunits, Δ6, Figure 4), GlxB (glyoxalase II, lack of which leads to accumulation of S-lactoyl-BSH, Δ27), YraH (putative glyoxalase, Δ3A), YdeA (Δ8) and YraA (Δ3A) (forming glyoxalase III that converts MG to D-lactate), YosR and BdbA (two disulfide oxidoreductases, Δ33), MsrA and MsrB (thiol stress-induced peptide methionine sulfoxide reductase) and the synthetic silencing of *bstG* (encoding a bacillithiol S-transferase, Δ8, Figure 3D; Supplementary Figure S7D). Overall, endogenous oxidative stress likely results from the deregulation and loss of the cell ability to sense and counter oxidative stress. However, this response appears to be complex, in particular because of a partial activation of the Spx-controlled regulon, which is limited to genes responding to thiol-oxidative stress (i.e. cysteine, Trx and BSH biosynthesis pathways; (50,51)) but does not include many Spx-controlled genes responding to disulfide stress, heat and cell wall stress (52).

In our genome-reduced strains, we found an astonishing level of resistance to the DNA damaging agent mitomycin C (MMC). Indeed, MGP254 is 300-fold more resistant to MMC than MS (Figure 5). MMC resistance varied

Table 3. Growth rates and biomass yields of evolved and reengineered strains

Medium	Strain	Genotype	Growth rate (h^{-1})	OD_{600nm} (AU)
NMS	MS	*trpC2* Δ*ICEBs1*, Δ*PBSX*, Δ*SPB*, Δ*skin*, Δ(*gtaB yvzH ggaB ggaA yvzI yvzE*), Δ*upp*::[λ*pR-neo*]	1.50 ± 0.05	1.85 ± 0.18
	MS^{evolA}	MS $rplQ_{G343A}$, $srfAA_{C8913T}$, $sigB_{G451A}$, $nhaC_{\Delta A17}$, $hutG_{\Delta G891}$	1.60 ± 0.09	2.31 ± 0.08
	MS^{evolB}	MS $rsbW_{C70G}$, $nhaC_{C515T}$, $ytcJ_{G294A}$, $yttP_{C316T}$, $hutG_{T323G}$	1.71 ± 0.12	2.49 ± 0.11
	$MS_{sigBG451A}$	MS $sigB_{G451A}$	1.48 ± 0.03	2.09 ± 0.02
	$MS_{rsbWC70G}$	MS $rsbW_{C70G}$	1.54 ± 0.11	2.07 ± 0.14
	$MS_{nhaC\Delta A17}$	MS $nhaC_{\Delta A17}$	1.71 ± 0.11	2.26 ± 0.10
	$MS_{nhaCC515T}$	MS $nhaC_{C515T}$	1.72 ± 0.04	2.27 ± 0.12
	$MS_{hutG\Delta G891}$	MS $hutG_{\Delta G891}$	1.61 ± 0.02	2.10 ± 0.10
	$MS_{hutGT323G}$	MS $hutG_{T323G}$	1.52 ± 0.10	1.99 ± 0.02
NMS	$MGP192^{p/o}$	MS_{cat} Δ(1, 2B, 3A, 4, 6, 7, 8, 9, 11, 13, 15, 17, 19, 23, 27, 29, 30, 33, 37, 38, 40, 42)$^{p/o}$	0.96 ± 0.01	2.11 ± 0.02
	$MGP229^{p/o}$	MS_{cat} Δ(1, 2B, 3A, 4, 6, 7, 8, 9, 11, 13, 15, 17, 19, 23, 27, 29, 30, 33, 37, 38, 40, 42, 43, 46)$^{p/o}$	1.01 ± 0.02	2.27 ± 0.01
	$MGP192^{p/o}{}_{nhaC\Delta A17}$	$MGP192^{p/o}nhaC_{\Delta A17}$	0.99 ± 0.02	2.19 ± 0.04
	$MGP192^{p/o}{}_{nhaCC515T}$	$MGP192^{p/o}nhaC_{C515T}$	0.90 ± 0.05	2.17 ± 0.00
	$MGP192^{p/o}{}_{sigBG451A}$	$MGP192^{p/o}sigB_{G451A}$	1.01 ± 0.06	2.14 ± 0.05
	$MGP192^{p/o}{}_{rsbWC70G}$	$MGP192^{p/o}rsbW_{C70G}$	0.96 ± 0.02	2.13 ± 0.04
	$MGP229^{p/o}{}_{nhaC\Delta A17}$	$MGP229^{p/o}nhaC_{\Delta A17}$	1.05 ± 0.02	2.28 ± 0.01
	$MGP229^{p/o}{}_{nhaCC515T}$	$MGP229^{p/o}nhaC_{C515T}$	1.07 ± 0.01	2.25 ± 0.00
	$MGP229^{p/o}{}_{rsbWC70G}$	$MGP229^{p/o}rsbW_{C70G}$	1.03 ± 0.04	2.29 ± 0.00
LB	$MGP229^{p/o}{}_{mutS}$	$MGP229^{p/o}$ Δ*mutS::kan*	0.77 ± 0.06	0.68 ± 0.08
	$MGP229^{p/o}{}_{mutSevolA}$	$MGP229^{p/o}$ Δ*mutS::kan* (+ mutations Table S8)	1.05 ± 0.07	0.80 ± 0.06
	$MGP229^{p/o}{}_{mutSevolB}$	$MGP229^{p/o}$ Δ*mutS::kan* (+ mutations Table S8)	1.02 ± 0.02	0.77 ± 0.06

across the lineage, with a logarithmic increase of resistance for the first 10 deletions combined in the genome, then a sharp loss of resistance followed by a sharp rise to maximum resistance. This profile suggests that the level of resistance to MMC is determined by multiple cellular functions. A component of increased MMC resistance may be the observed upregulation of the LexA regulon, which responds to DNA damage (45). In genome-reduced strains, LexA is activated constitutively without exposure to external DNA damaging agents (Figure 3). As all the genes known to participate in DNA recombination and repair processes are present, increased expression of DNA repair machineries may increase cell resistance to MMC lesions. Another factor contributing to MMC resistance may be the removal of membrane-associated proteins such as transporters, which could limit the uptake of MMC in the cell. However, these two factors are not sufficient to explain the observed recovery of sensitivity to MMC for strains in the middle of the lineage (Figure 5), and future investigations will be required to characterize mechanisms underlying this recovery. A third factor contributing to MMC resistance may be related to the response to thiol-oxidative stress observed in genome-reduced cells (Figure 4). Indeed, MMC reacts with thiols to undergo reductive activation and form covalent adducts with DNA (53). Additional pathways for the modification of MMC by dithiols were identified (54), raising the possibility that MMC could be further modified and inactivated in the intracellular environment of genome-reduced cells. Activated MMC was also shown to covalently modify the active site of thioredoxin reductase *in vitro* and in cancer cells (55,56), raising the possibility that the observed TrxB upregulation may contribute to the MMC resistance of the genome-reduced strains.

A second remarkable emerging phenotype, which could not be predicted from known molecular pathways, is the drastic reduction of spontaneous mutagenesis observed in our genome-reduced strains (Table 2). The frequency of spontaneous mutations in the *rpoB* gene conferring resistance to rifampicin was 100-fold lower in MGP229 that in MS. This finding was surprising because, in *E. coli,* genome reduction was reported to increase mutation rate (15). A factor contributing to the low spontaneous mutagenesis may be the observed overexpression of *uvrAB* encoding a key component of the nucleotide excision repair system, which recognizes damaged DNA bases. Other factors may involve the interplay between membrane-associated proteins and methyl-directed mismatch repair (MMR) system. In *E. coli*, expression of the multidrug efflux pump AcrAB negatively correlates with *mutS* expression and positively correlates with the mutation frequency (57). Although neither *mutSL* nor *mutS2* were transcriptionally deregulated in the *B. subtilis* genome-reduced strains, we cannot exclude the possibility that the deletion of more than 140 genes annotated as '*transporters*' may affect MMR-dependent mutation frequencies.

Inactivation of *mutS* in genome-reduced strains restored the mutation rate to the level observed in the Δ*mutS* reference strain. We cannot exclude the existence of another yet unknown DNA repair mechanism, more active in genome-reduced strains than in the reference strain, which would be overloaded by mutations caused by the MMR inactivation. We also cannot exclude the possibility that the spontaneous mutation rate in genome-reduced strains is inherently lower than in wild-type strains, independent of the activity of any repair mechanism. Yet, the principle of parsimony leads us to propose that the low mutation rate in genome-reduced

strains depends on MMR activity. The high MMR activity and low spontaneous mutagenesis also explains the inability of genome-reduced strains to acquire beneficial mutations conferring fitness gains during short-term evolution experiments (Table 3). However, such beneficial mutations were readily selected during evolution experiments with MMR-deficient cells. The putative beneficial mutations we identified (Supplementary Table S8) likely corrected some of the transcriptional and metabolic deregulations accumulated during genome reduction, generating more competitive isolates. In combination with the loss of genetic competence observed in this work (Figure 3) and in a previous study (12), *B. subtilis* cells with reduced genomes appear to have a limited capacity to evolve spontaneously. Developing strains with reduced evolutionary potential may increase the genetic reliability of synthetic biology devices (58). Although low evolvability (i.e. low spontaneous mutagenesis and low genetic competence) emerged from genome reduction (rather than being engineered), this property combined with tools to transiently restore evolvability (e.g. induction of spontaneous mutagenesis or genetic competence) could lead to the future development of chassis strains minimizing evolutionary failure of synthetic devices. The development of an *E. coli* strain with reduced plasmid mutation rates by 6- to 30-fold (59), already proved that it is possible to improve the genetic stability of cellular chassis without encountering trade-offs in other desirable performance characteristics.

Synthetic cells are valuable platforms that can be used to discover yet undetected biological functions and shed light on the fundamental principles of the minimum requirements for sustaining life. Genome-reduced strains also help discover novel gene functions and regulatory mechanisms, establish improved whole-cell models, and possibly limit evolutionary failure of synthetic functions. Because *B. subtilis* is widely used in industry, the development of fitness-improved custom-designed genomes is a major challenge to meet the demand for novel biotechnological applications. Minimal genomes are not per se a prerequisite for chassis construction albeit they represent an attractive alternative for precise control over regulatory and metabolic programs. For instance, the generation of libraries of engineered microorganisms significantly different from those found in nature and dedicated to specific applications operating as single species or in communities represents the next generation of bacterial chassis. Although our work points out the complexity of *B. subtilis* metabolic and regulatory networks and our limited knowledge of these networks, it also reveals two novel, seemingly independent emergent properties; (i) robustness to DNA damaging agent, possibly associated with a constitutive thiol-oxidative stress, and (ii) low spontaneous mutagenesis, possibly dependent on high MMR activity. The mechanisms underlying these emerging properties will need further characterization to be able to rationally develop new chassis in which these properties are engineered (e.g. low evolvability). Thus, design and construction of efficient cell factories shall be envisioned as an engineering framework coupling reductive and evolutive cycles with high-throughput screening. Reductive cycles will encompass successive deletions of unwanted genomic regions in a context of low evolvability. Evolutive cycles will include a condition-dependent increase of spontaneous mutagenesis for experimental evolution or the integration in a chromosomal landing-pad of a library of DNA fragments from the *Bacillus* species pangenome. This blueprint will allow selecting efficient cell factories at each cycle of deletion, before entering the next cycle and pave the way for innovative construction of synthetic chassis.

DATA AVAILABILITY

Gene expression data have been deposited with the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE207089 and can be visualized with the Genoscapist *B. subtilis* Expression Data Browser (http://genoscapistdev.migale.inrae.fr/smgeb/).

Genoscapist is an open-source initiative available in the GitHub repository (https://github.com/sderozier/genoscapist) and in the Zenodo repository (https://zenodo.org/record/7645915).

Genome sequences of 43 genome-reduced strains are available in the NCBI Database at https://www.ncbi.nlm.nih.gov/nuccore/ and can be accessed with GenBank accession numbers from CP116828 to CP116870.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We acknowledge the Migale platform (INRAE, Jouy-en-Josas, France) for their support for providing useful software for genome sequence analysis.

Author contributions: M.J., P.N., K-I.Y., U.S. and P.Ni. conceived the project and designed the experimental plan, E.D., A-G.P., K.T., V.C. and F.L. performed the experiments, M.J., C.G. and P.Ni. performed statistical treatments, M.J., E.D. and A-G.P. analyzed the data, C.G. and S.D. extended the genoscapist genome browser, M.J., P.N., E.D. and A-G.P. wrote the manuscript.

FUNDING

French National Research Agency [ANR-18-CE44-0003, ANR-18-CE43-0002, ANR-15-CE21-0008]; European Commission [LSHG-CT-2006-037469, FP7-244093]. Funding for open access charge: French National Research Agency [ANR-18-CE43-0002, ANR-18-CE44-0003].

Conflict of interest statement. None declared.

REFERENCES

1. Pál,C., Papp,B. and Pósfai,G. (2014) The dawn of evolutionary genome engineering. *Nat. Rev. Genet.*, **15**, 504–512.
2. Wang,H.H., Isaacs,F.J., Carr,P.A., Sun,Z.Z., Xu,G., Forest,C.R. and Church,G.M. (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, **460**, 894–898.
3. Isaacs,F.J., Carr,P.A., Wang,H.H., Lajoie,M.J., Sterling,B., Kraal,L., Tolonen,A.C., Gianoulis,T.A., Goodman,D.B., Reppas,N.B. *et al.* (2011) Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science*, **333**, 348–353.

4. Lajoie,M.J., Rovner,A.J., Goodman,D.B., Aerni,H.R., Haimovich,A.D., Kuznetsov,G., Mercer,J.A., Wang,H.H., Carr,P.A., Mosberg,J.A. *et al.* (2013) Genomically recoded organisms expand biological functions. *Science*, **342**, 357–360.
5. Hutchison,C.A. 3rd, Chuang,R.Y., Noskov,V.N., Assad-Garcia,N., Deerinck,T.J., Ellisman,M.H., Gill,J., Kannan,K., Karas,B.J., Ma,L. *et al.* (2016) Design and synthesis of a minimal bacterial genome. *Science*, **351**, aad6253.
6. Gibson,D.G., Glass,J.I., Lartigue,C., Noskov,V.N., Chuang,R.Y., Algire,M.A., Benders,G.A., Montague,M.G., Ma,L., Moodie,M.M. *et al.* (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*, **329**, 52–56.
7. Li,Y., Zhu,X., Zhang,X., Fu,J., Wang,Z., Chen,T. and Zhao,X. (2016) Characterization of genome-reduced *Bacillus subtilis* strains and their application for the production of guanosine and thymidine. *Microb. Cell Fact.*, **15**, 94.
8. Manabe,K., Kageyama,Y., Morimoto,T., Ozawa,T., Sawada,K., Endo,K., Tohata,M., Ara,K., Ozaki,K. and Ogasawara,N. (2011) Combined effect of improved cell yield and increased specific productivity enhances recombinant enzyme production in genome-reduced *Bacillus subtilis* strain MGB874. *Appl. Environ. Microbiol.*, **77**, 8370–8381.
9. Reuss,D.R., Commichau,F.M., Gundlach,J., Zhu,B. and Stulke,J. (2016) The Blueprint of a Minimal Cell: miniBacillus. *Microbiol. Mol. Biol. Rev.*, **80**, 955–987.
10. Tanaka,K., Henry,C.S., Zinner,J.F., Jolivet,E., Cohoon,M.P., Xia,F., Bidnenko,V., Ehrlich,S.D., Stevens,R.L. and Noirot,P. (2013) Building the repertoire of dispensable chromosome regions in *Bacillus subtilis* entails major refinement of cognate large-scale metabolic model. *Nucleic Acids Res.*, **41**, 687–699.
11. Kurokawa,M., Seno,S., Matsuda,H. and Ying,B.-W. (2016) Correlation between genome reduction and bacterial growth. *DNA Res.*, **23**, 517–525.
12. Reuss,D.R., Altenbuchner,J., Mader,U., Rath,H., Ischebeck,T., Sappa,P.K., Thurmer,A., Guerin,C., Nicolas,P., Steil,L. *et al.* (2017) Large-scale reduction of the *Bacillus subtilis* genome: consequences for the transcriptional network, resource allocation, and metabolism. *Genome Res.*, **27**, 289–299.
13. Choe,D., Lee,J.H., Yoo,M., Hwang,S., Sung,B.H., Cho,S., Palsson,B., Kim,S.C. and Cho,B.-K. (2019) Adaptive laboratory evolution of a genome-reduced *Escherichia coli*. *Nat. Commun.*, **10**, 935.
14. Karcagi,I., Draskovits,G., Umenhoffer,K., Fekete,G., Kovács,K., Méhi,O., Balikó,G., Szappanos,B., Györfy,Z., Fehér,T. *et al.* (2016) Indispensability of horizontally transferred genes and its impact on bacterial genome streamlining. *Mol. Biol. Evol.*, **33**, 1257–1269.
15. Nishimura,I., Kurokawa,M., Liu,L. and Ying,B.-W. (2017) Coordinated changes in mutation and growth rates induced by genome reduction. *MBio*, **8**, e00676–e00617.
16. Dillon,M.M., Sung,W., Lynch,M. and Cooper,V.S. (2018) Periodic variation of mutation rates in bacterial genomes associated with replication timing. *MBio*, **9**, e01371-18.
17. Bjedov,I., Tenaillon,O., Gérard,B., Souza,V., Denamur,E., Radman,M., Taddei,F. and Matic,I. (2003) Stress-induced mutagenesis in bacteria. *Science*, **300**, 1404–1409.
18. Krašovec,R., Belavkin,R.V., Aston,J.A.D., Channon,A., Aston,E., Rash,B.M., Kadirvel,M., Forbes,S. and Knight,C.G. (2014) Mutation rate plasticity in rifampicin resistance depends on *Escherichia coli* cell–cell interactions. *Nat. Commun.*, **5**, 3742.
19. Wrande,M., Roth,J.R. and Hughes,D. (2008) Accumulation of mutants in "aging" bacterial colonies is due to growth under selection, not stress-induced mutagenesis. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 11863–11868.
20. Kawecki,T.J., Lenski,R.E., Ebert,D., Hollis,B., Olivieri,I. and Whitlock,M.C. (2012) Experimental evolution. *Trends Ecol. Evol.*, **27**, 547–560.
21. Barrick,J.E. and Lenski,R.E. (2013) Genome dynamics during experimental evolution. *Nat. Rev. Genet.*, **14**, 827–839.
22. Kobayashi,K., Ehrlich,S.D., Albertini,A., Amati,G., Andersen,K.K., Arnaud,M., Asai,K., Ashikaga,S., Aymerich,S., Bessieres,P. *et al.* (2003) Essential *Bacillus subtilis* genes. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 4678–4683.
23. Koo,B.M., Kritikos,G., Farelli,J.D., Todor,H., Tong,K., Kimsey,H., Wapinski,I., Galardini,M., Cabal,A., Peters,J.M. *et al.* (2017) Construction and analysis of two genome-scale deletion libraries for *Bacillus subtilis*. *Cell Syst.*, **4**, 291–305.
24. Borkowski,O., Goelzer,A., Schaffer,M., Calabre,M., Mader,U., Aymerich,S., Jules,M. and Fromion,V. (2016) Translation elicits a growth rate-dependent, genome-wide, differential protein production in *Bacillus subtilis*. *Mol. Syst. Biol.*, **12**, 870.
25. Buescher,J.M., Liebermeister,W., Jules,M., Uhr,M., Muntel,J., Botella,E., Hessling,B., Kleijn,R.J., Le Chat,L., Lecointe,F. *et al.* (2012) Global network reorganization during dynamic adaptations of *Bacillus subtilis* metabolism. *Science*, **335**, 1099–1103.
26. Guiziou,S., Sauveplane,V., Chang,H.J., Clerte,C., Declerck,N., Jules,M. and Bonnet,J. (2016) A part toolbox to tune genetic expression in *Bacillus subtilis*. *Nucleic Acids Res.*, **44**, 7495–7508.
27. Kunst,F., Ogasawara,N., Moszer,I., Albertini,A.M., Alloni,G., Azevedo,V., Bertero,M.G., Bessieres,P., Bolotin,A., Borchert,S. *et al.* (1997) The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature*, **390**, 249–256.
28. Nicolas,P., Mäder,U., Dervyn,E., Rochat,T., Leduc,A., Pigeonneau,N., Bidnenko,E., Marchadier,E., Hoebeke,M., Aymerich,S. *et al.* (2012) Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*. *Science*, **335**, 1103–1106.
29. Anagnostopoulos,C. and Spizizen,J. (1961) Requirements for transformation in *Bacillus subtilis*. *J. Bacteriol.*, **81**, 741–746.
30. Westers,H., Dorenbos,R., van Dijl,J.M., Kabel,J., Flanagan,T., Devine,K.M., Jude,F., Seror,S.J., Beekman,A.C., Darmon,E. *et al.* (2003) Genome engineering reveals large dispensable regions in *Bacillus subtilis*. *Mol. Biol. Evol.*, **20**, 2076–2090.
31. Auchtung,J.M., Lee,C.A., Monson,R.E., Lehman,A.P. and Grossman,A.D. (2005) Regulation of a *Bacillus subtilis* mobile genetic element by intercellular signaling and the global DNA damage response. *Proc. Nat. Acad. Sci. U.S.A.*, **102**, 12554–12559.
32. Fabret,C., Ehrlich,S.D. and Noirot,P. (2002) A new mutation delivery system for genome-scale approaches in *Bacillus subtilis*. *Mol. Microbiol.*, **46**, 25–36.
33. Derozier,S., Nicolas,P., Mader,U. and Guerin,C. (2021) Genoscapist: online exploration of quantitative profiles along genomes via interactively customized graphical representations. *Bioinformatics*, **37**, 2747–2749.
34. Henry,C.S., Zinner,J.F., Cohoon,M.P. and Stevens,R.L. (2009) i*Bsu*1103: a new genome-scale metabolic model of *Bacillus subtilis* based on SEED annotations. *Genome Biol.*, **10**, R69.
35. Zhu,B. and Stülke,J. (2017) SubtiWiki in 2018: from genes and proteins to functional network annotation of the model organism *Bacillus subtilis*. *Nucleic Acids Res.*, **46**, D743–D748.
36. Gardan,R., Rapoport,G. and Débarbouillé,M. (1997) Role of the transcriptional activator RocR in the arginine-degradation pathway of *Bacillus subtilis*. *Mol. Microbiol.*, **24**, 825–837.
37. Commichau,F.M., Herzberg,C., Tripal,P., Valerius,O. and Stülke,J. (2007) A regulatory protein–protein interaction governs glutamate biosynthesis in *Bacillus subtilis*: the glutamate dehydrogenase RocG moonlights in controlling the transcription factor GltC. *Mol. Microbiol.*, **65**, 642–654.
38. Calvo,R.A. and Kearns,D.B. (2015) FlgM is secreted by the flagellar export apparatus in *Bacillus subtilis*. *J. Bacteriol.*, **197**, 81–91.
39. Courtney,C.R., Cozy,L.M. and Kearns,D.B. (2012) Molecular characterization of the flagellar hook in *Bacillus subtilis*. *J. Bacteriol.*, **194**, 4619–4629.
40. Lu,J. and Holmgren,A. (2014) The thioredoxin antioxidant system. *Free Radic Biol Med*, **66**, 75–87.
41. Zheng,C., Guo,S., Tennant,W.G., Pradhan,P.K., Black,K.A. and Dos Santos,P.C. (2019) The thioredoxin system reduces protein persulfide intermediates formed during the synthesis of thio-cofactors in *Bacillus subtilis*. *Biochemistry*, **58**, 1892–1904.
42. Guillouard,I., Auger,S., Hullo,M.-F., Chetouani,F., Danchin,A. and Martin-Verstraete,I. (2002) Identification of *Bacillus subtilis* CysL, a regulator of the *cysJI* operon, which encodes sulfite reductase. *J. Bacteriol.*, **184**, 4681–4689.
43. Gaballa,A., Newton,G.L., Antelmann,H., Parsonage,D., Upton,H., Rawat,M., Claiborne,A., Fahey,R.C. and Helmann,J.D. (2010) Biosynthesis and functions of bacillithiol, a major low-molecular-weight thiol in *Bacilli*. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 6482–6486.

44. Chandrangsu,P., Dusi,R., Hamilton,C.J. and Helmann,J.D. (2014) Methylglyoxal resistance in *Bacillus subtilis*: contributions of bacillithiol-dependent and independent pathways. *Mol. Microbiol.*, **91**, 706–715.
45. Au,N., Kuester-Schoeck,E., Mandava,V., Bothwell,L.E., Canny,S.P., Chachu,K., Colavito,S.A., Fuller,S.N., Groban,E.S., Hensley,L.A. *et al.* (2005) Genetic composition of the *Bacillus subtilis* SOS system. *J. Bacteriol.*, **187**, 7655–7666.
46. Duigou,S., Ehrlich,S.D., Noirot,P. and Noirot-Gros,M.-F. (2004) Distinctive genetic features exhibited by the Y-family DNA polymerases in *Bacillus subtilis*. *Mol. Microbiol.*, **54**, 439–451.
47. Iyer,R.R., Pluciennik,A., Burdett,V. and Modrich,P.L. (2006) DNA mismatch repair: functions and mechanisms. *Chem. Rev.*, **106**, 302–323.
48. Sung,W., Ackerman,M.S., Gout,J.-F., Miller,S.F., Williams,E., Foster,P.L. and Lynch,M. (2015) Asymmetric context-dependent mutation patterns revealed through mutation–accumulation experiments. *Mol. Biol. Evol.*, **32**, 1672–1683.
49. Brito,P.H., Chevreux,B., Serra,C.R., Schyns,G., Henriques,A.O. and Pereira-Leal,J.B. (2017) Genetic competence drives genome diversity in *Bacillus subtilis*. *Genome Biol. Evol.*, **10**, 108–124.
50. Nakano,S., Küster-Schöck,E., Grossman,A.D. and Zuber,P. (2003) Spx-dependent global transcriptional control is induced by thiol-specific oxidative stress in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 13603–13608.
51. Rochat,T., Nicolas,P., Delumeau,O., Rabatinová,A., Korelusová,J., Leduc,A., Bessières,P., Dervyn,E., Krásný,L. and Noirot,P. (2012) Genome-wide identification of genes directly regulated by the pleiotropic transcription factor Spx in *Bacillus subtilis*. *Nucleic Acids Res.*, **40**, 9571–9583.
52. Rojas-Tapias,D.F. and Helmann,J.D. (2019) Roles and regulation of Spx family transcription factors in *Bacillus subtilis* and related species. *Adv. Microb. Physiol.*, **75**, 279–323.
53. Paz,M.M. (2009) Reductive activation of mitomycin C by thiols: kinetics, mechanism, and biological implications. *Chem. Res. Toxicol.*, **22**, 1663–1668.
54. Paz,M.M. (2010) Cross-linking of dithiols by mitomycin C. *Chem. Res. Toxicol.*, **23**, 1384–1392.
55. Paz,M.M., Zhang,X., Lu,J. and Holmgren,A. (2012) A new mechanism of action for the anticancer drug mitomycin C: mechanism-based inhibition of thioredoxin reductase. *Chem. Res. Toxicol.*, **25**, 1502–1511.
56. Yokomizo,A., Ono,M., Nanri,H., Makino,Y., Ohga,T., Wada,M., Okamoto,T., Yodoi,J., Kuwano,M. and Kohno,K. (1995) Cellular levels of thioredoxin associated with drug sensitivity to cisplatin, mitomycin C, doxorubicin, and etoposide. *Cancer Res.*, **55**, 4293–4296.
57. El Meouche,I. and Dunlop,M.J. (2018) Heterogeneity in efflux pump expression predisposes antibiotic-resistant cells to mutation. *Science*, **362**, 686–690.
58. Renda,B.A., Hammerling,M.J. and Barrick,J.E. (2014) Engineering reduced evolutionary potential for synthetic biology. *Mol. Biosyst.*, **10**, 1668–1678.
59. Deatherage,D.E., Leon,D., Rodriguez,A.E., Omar,S.K. and Barrick,J.E. (2018) Directed evolution of *Escherichia coli* with lower-than-natural plasmid mutation rates. *Nucleic Acids Res.*, **46**, 9236–9250.