ARTICLES
https://doi.org/10.1038/s41559-019-0939-6
nature ecology & evolution

# Microbial life cycles link global modularity in regulation to mosaic evolution

Jordi van Gestel[1,2,3,4]*, Martin Ackermann[3,4] and Andreas Wagner[1,2,5]*

**Microbes are exposed to changing environments, to which they can respond by adopting various lifestyles such as swimming, colony formation or dormancy. These lifestyles are often studied in isolation, thereby giving a fragmented view of the life cycle as a whole. Here, we study lifestyles in the context of this whole. We first use machine learning to reconstruct the expression changes underlying life cycle progression in the bacterium *Bacillus subtilis*, based on hundreds of previously acquired expression profiles. This yields a timeline that reveals the modular organization of the life cycle. By analysing over 380 Bacillales genomes, we then show that life cycle modularity gives rise to mosaic evolution in which life stages such as motility and sporulation are conserved and lost as discrete units. We postulate that this mosaic conservation pattern results from habitat changes that make these life stages obsolete or detrimental. Indeed, when evolving eight distinct Bacillales strains and species under laboratory conditions that favour colony growth, we observe rapid and parallel losses of the sporulation life stage across species, induced by mutations that affect the same global regulator. We conclude that a life cycle perspective is pivotal to understanding the causes and consequences of modularity in both regulation and evolution.**

Microbes express an incredible range of lifestyles, from the myriad of planktonic life forms floating in the oceans to the sessile biofilms sticking to our teeth[1–5]. In many cases, cells adjust their lifestyle in response to specific environmental conditions[4,6,7]. They can 'choose' to swim, to form colonies or to become dormant depending on these conditions[8]. When studied in the context of a life cycle, these flexible lifestyles are often referred to as life stages through which cells transition in time[3,9–11]. Over recent decades, lifestyle transitions have been characterized in great detail by deciphering the gene regulatory networks underlying cellular decision making[12–15]. Most of these studies focus on the role of a few specific regulators underlying one of the lifestyle transitions, in part because it is challenging to monitor the hundreds of genes underlying life cycle progression across the entire life cycle of a microbe. Here, we overcome this challenge in *Bacillus subtilis* by employing a machine learning approach to reconstruct the expression changes along its complete life cycle. We subsequently use this reconstruction to study both the organization and evolution of life cycles across the phylogenetic order of the Bacillales.

*Bacillus subtilis* can adopt a variety of lifestyles (Fig. 1). First, its cells can swim either individually or through swarming motility as part of a group[16,17]. Second, cells can form sessile colonies, which are typically called biofilms[18–20]. These colonies can, for example, arise on plant roots, where cells consume root exudates to produce the extracellular polysaccharides that hold the colony together[21]. Third and finally, cells can enter dormancy by triggering the endosporulation process[12]. During sporulation, cells undergo asymmetric cell division in which the smaller (daughter) cell is engulfed by the larger (mother) cell and develops into a spore. Spores are released into the environment following lysis of the mother cell. They can resist extreme environmental stresses[22] while being spring-loaded with life[23], ready to germinate whenever conditions improve.

We start our analysis by synthesizing data from previous studies on the global transcription network of *B. subtilis*, with a particular focus on the global regulatory modules underlying lifestyle transitions. Then, we reconstruct the life cycle of *B. subtilis* using auto-associative artificial neural networks based on hundreds of previously acquired gene expression profiles. This analysis provides a detailed timeline of the expression of genes across the *B. subtilis* life cycle. We use this timeline to examine the evolution of Bacillales life cycles, through both comparative genomics and experimental evolution of Bacillales strains and species. Our results demonstrate the key importance of the life cycle in linking regulatory modularity to mosaic evolution.

## Results

**The global transcription network has a simple organization, shaped by a few global regulators and complementary regulatory modules.** Before examining how different lifestyles are embedded in the life cycle of *B. subtilis*, we first briefly examine their regulatory underpinnings by reviewing the global transcription network of *B. subtilis*[12,18,24–34]; specifically, we compare four reconstructions of its transcription network. These reconstructions were made as part of three independent studies[28,34,35], in an attempt to match the real transcription network of *B. subtilis* as closely as possible (Supplementary Text 1). Even though the studies differed in respect to both the timing and methodology by which the network was reconstructed, their reconstructions share many similarities (Supplementary Fig. 1). First, the majority of regulatory interactions are controlled by a small number of transcriptional regulators[32]. Depending on the specific reconstruction, between seven and 17 regulators control most regulatory interactions inside the cell. We refer to these regulators as 'global regulators'. Second, consistent with findings in *Escherichia coli* and *Saccharomyces cerevisiae*[36–39], there are many regulatory interactions among the global regulators

[1]Department of Evolutionary Biology and Environmental Studies, University of Zürich, Zürich, Switzerland. [2]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [3]Department of Environmental Systems Science, ETH Zürich, Zürich, Switzerland. [4]Department of Environmental Microbiology, Swiss Federal Institute of Aquatic Science and Technology (Eawag), Dübendorf, Switzerland. [5]The Santa Fe Institute, Santa Fe, NM, USA. *e-mail: jordivangestel@gmail.com; andreas.wagner@ieu.uzh.ch

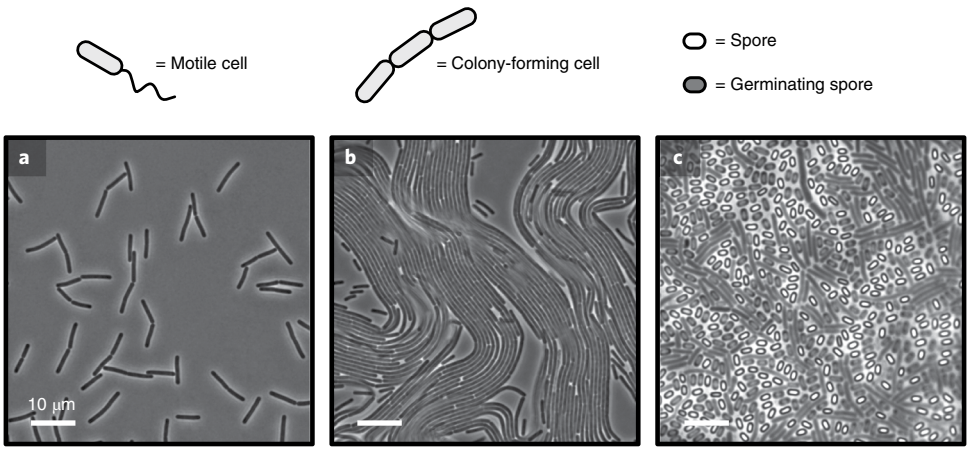


**Fig. 1 | Lifestyles of *B. subtilis*.** Examples of different lifestyle of *B. subtilis*. **a**, Motile lifestyle (flagella not visible). **b**, Colony-forming lifestyle, in which cells secrete extracellular polysaccharides and amyloid fibres that result in bundle formation[20]. **c**, Dormant lifestyle, in which cells form phase-bright spores. Some of these spores are germinating, shown by the transition from a phase-bright to a phase-dark spore.

whereas regulatory interactions in general are sparse. Hence, global regulators do not act independently but act together in affecting the global expression state of a cell. Third, global regulators consist of both transcription factors and sigma factors, of which the latter account for most interactions[40–42]. Sigma factors are part of the RNA polymerase holo-enzyme and are essential for the initiation of transcription in bacteria[43].

Given that global regulators control the expression of most genes, the question emerges of whether they control complementary or overlapping sets of genes. To address this question, we compare the regulons of the top 15 global regulators—that is, the regulators that regulate the most genes—in a pairwise fashion. We detail our observations for the most recent reconstruction of the global transcription network[35], but obtain similar results for other reconstructions (Supplementary Fig. 2). The circles in Fig. 2 show the expected and observed overlap between regulons for each pairwise combination of regulators. The expectation is based on the null hypothesis that global regulators affect a random set of genes of the same size as the observed regulon, while ignoring regulatory structures such as operons (see Methods). When the overlap between two regulons is higher than expected, a circle is coloured red and when it is lower, a circle is coloured blue. Figure 2 reveals a strikingly simple global organization, in which most regulons show little to no overlap (that is, most circles are blue) but where five modules of regulons overlap strongly (see black squares). These modules correspond to distinct sets of previously identified cellular functions[32,35], including the stress response, metabolism and—especially relevant here—the different lifestyles of *B. subtilis*, with many global regulators regulating sporulation. In sum, the global expression state of a cell depends on a small number of global regulators, which regulate each other and a small number of downstream functional programmes.

**The timeline of the *B. subtilis* life cycle can be reconstructed from expression data**. To study how the lifestyles of *B. subtilis* are embedded in its overall life cycle, we investigate how regulators and their downstream targets are expressed over time. Experimental studies broadly use two different approaches to acquire such data. The first employs fluorescent reporters to monitor the expression of genes by time-lapse microscopy[14,15,44,45]. This approach allows for data acquisition at high temporal resolution, but is limited by the number of genes and environments that can be monitored. The second acquires transcriptomic data, which yield information about hundreds of genes simultaneously but usually for a limited number of time points[46–48]. Inspired by 'pseudo-time' reconstruction in developmental biology[49–54], we developed a third approach: we use machine learning to reconstruct the expression changes underlying life cycle progression from many previously acquired gene expression profiles.

The rationale behind life cycle reconstruction is that global expression changes occur gradually in time, and can therefore be reconstructed by sorting gene expression profiles according to their similarity along a single dimension, even though such profiles may have been acquired from very different environments (see Methods). To achieve this reconstruction, we use the largest transcriptomic analysis that is currently available for *B. subtilis*. This used tiling micro-arrays with a resolution of 22 base pairs to examine the expression of *B. subtilis* 168 under 104 different environmental conditions[42]. These environments are highly diverse: they include minimal and rich media, low and high temperatures, liquid and solid media, harsh and benign conditions and germination and sporulation conditions. Of the 269 gene expression profiles in this study[42], we used those 252 derived from the same genetic variant (see Methods). For our analysis, we specifically focus on 'lifestyle' genes, which we define as genes in the core regulatory cascades underlying lifestyle transitions, including their associated regulatory targets. For our purpose, it is not important to identify all such genes. Instead, we focus on genes that (1) have repeatedly been associated with lifestyle biology in *B. subtilis*, (2) are extensively studied in the literature[8,18,19,22,31,35,55–57] and (3) are part of the global transcription network discussed above. Supplementary Fig. 3 gives an overview of the genes included in our analysis and of the regulatory interactions among them. Importantly, our results do not rely on the specific selection of genes, because the same qualitative results are obtained for alternative random selections of lifestyle genes (see Methods and Supplementary Fig. 7).

To sort the expression profiles of the lifestyle genes along a single dimension, we use auto-associative artificial neural networks[58–60]. Such networks are central to a machine learning approach that maps a multi-dimensional dataset onto a single dimension (see Methods and Supplementary Text 2). In this approach, a neural network is trained such that the input—here, the gene expression profile of lifestyle genes—matches the output as closely as possible. Importantly, the network employed in this approach harbours an information bottleneck in the form of a single node through which all information must pass (Supplementary Fig. 5). This forces the network to reduce multi-dimensional gene expression profiles onto a single dimension, which effectively sorts such profiles according to their similarity in this dimension[59,60] (see Methods). Figure 3a shows the

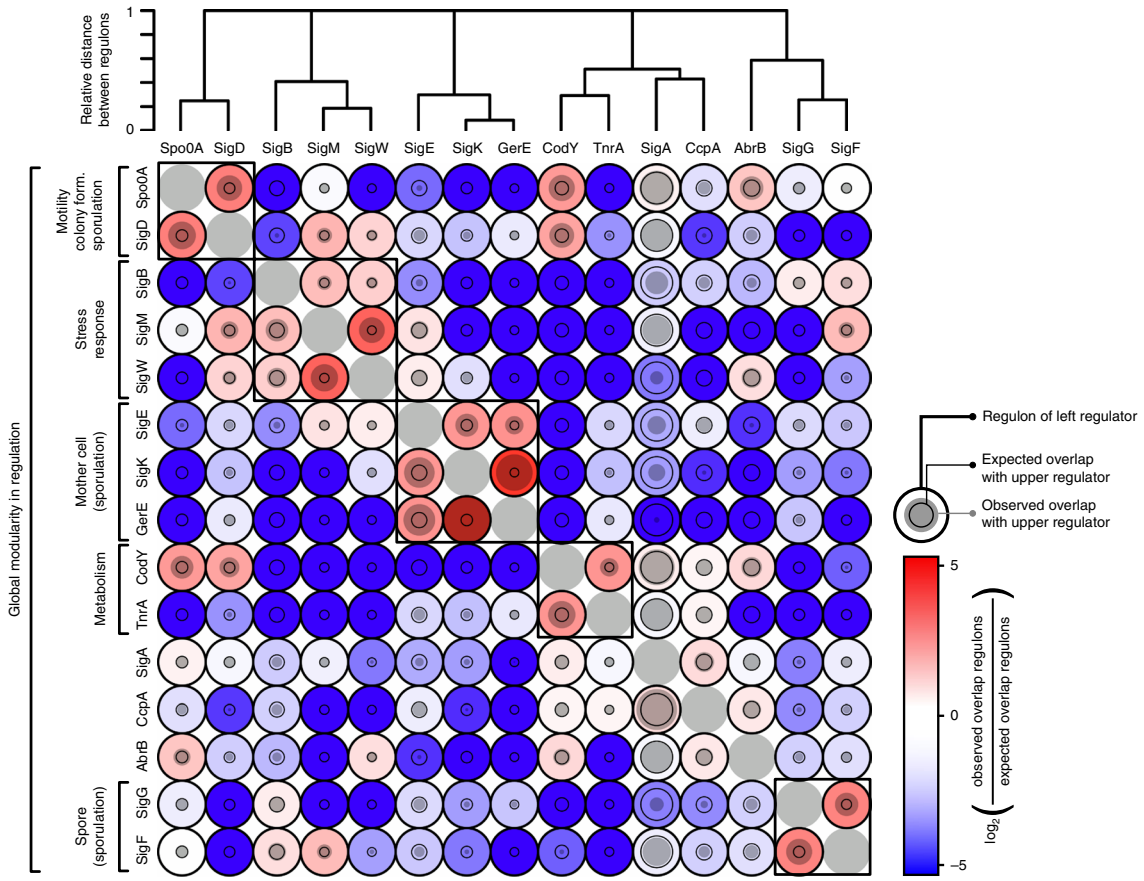


**Fig. 2 | Global transcriptional regulation shows a modular organization.** The overlap between regulons of each pairwise combination of 15 global regulators is illustrated by three circles: the outer circle represents the regulon of the global regulator shown on the left (for simplicity, all outer circles are of the same size even though they represent different regulon sizes); the inner circle represents the fraction of this regulon that is expected to be coregulated with the upper global regulator; and the grey circle shows the observed fraction that is coregulated. When the observed fraction (that is, the grey circle) is smaller/greater than the expected fraction (that is, the inner circle), the background of the outer circle is shown in blue/red. A hierarchical clustering algorithm to detect higher-level regulatory modules among the regulons of global regulators reveals five clearly distinct modules that are outlined by black squares. Functional categories of the global regulators are based on the functional annotation of the SubtiWiki database[35,137] (Supplementary Text 1).

results of training a single neural network on our expression dataset. Each blue data point corresponds to one of the 252 expression profiles. The straight red-lined segments indicate how these profiles are mapped onto the single dimension (curved red line). Most expression profiles align closely to this single dimension: when training 100 independent neural networks (Fig. 3b), only ten of the 252 expression profiles cannot robustly be mapped onto the single dimension (Supplementary Fig. 6). We excluded these ten ambiguous profiles, corresponding to three environmental conditions, in subsequent analyses (see Methods).

In principle, the single dimension that is the outcome of this dimensionality reduction method could correspond to any environmental variable. Nevertheless, in line with our expectations, it turns out that this variable relates to time and captures the temporal expression changes along the life cycle of *B. subtilis*. First, expression profiles at one end of the single dimension are consistently associated with germination conditions, while those at the opposite end are associated with sporulation conditions (Fig. 3a). Second, for time series expression data shown in ref. [42], we observe a strong correlation between the experimental time at which an expression profile is acquired and the projection of the profile onto the single dimension (Supplementary Figs. 10 and 11). Thus, the dimension can best be interpreted as pseudo-time, a relative time along the life cycle of *B. subtilis*, from the moment of germination to sporulation. Our interpretation of pseudo-time corresponds to that in developmental biology, where sorted expression profiles are often used to infer temporal expression changes in both cell differentiation and embryogenesis[49–54]. By the same token, the pseudo-time dimension gives us the unique opportunity to examine all expression changes over the life cycle of *B. subtilis*.

**The life cycle of *B. subtilis* has a modular organization, guided by the serial expression of global regulators.** Our observation that most regulatory interactions are controlled by only a few global regulators (Fig. 2) led us to the next hypothesis: the expression changes in pseudo-time result from the serial expression of global regulators and their downstream regulatory modules. To test this hypothesis, we first focus on effector genes underlying the distinct lifestyles. Effector genes bring forth the phenotypic properties of a cell and are not involved in regulation. In consequence, they have fewer pleiotropic effects than regulatory genes. We divide the lifestyle effector genes into three categories (Supplementary Table 3): (1) those underlying motility (for example, flagella formation, chemotaxis); (2) those underlying colony formation and associated starvation phenotypes, such as antimicrobial production, surfactin production and competence; and (3) those underlying sporulation and

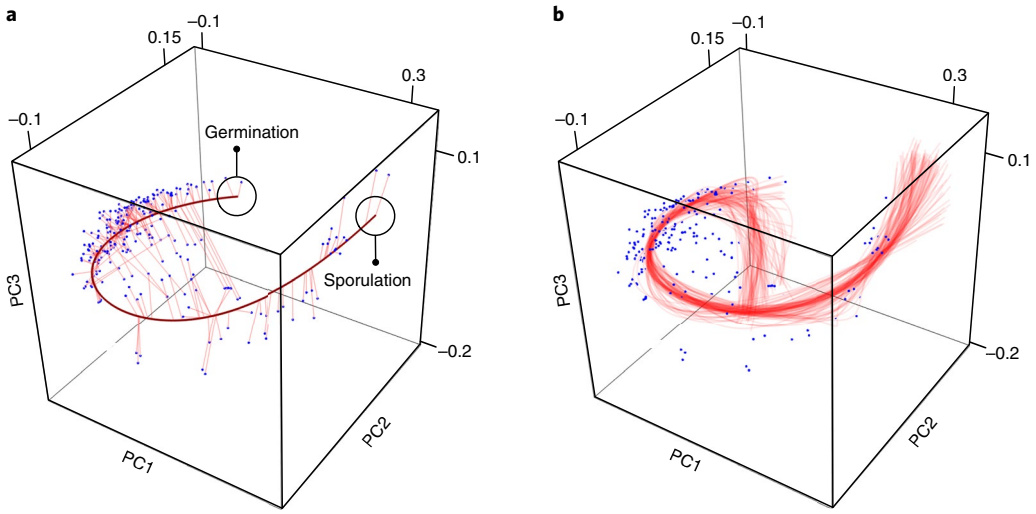


**Fig. 3 | Expression profiles can be sorted along a single (pseudo-time) dimension. a**, A total of 252 expression profiles (blue dots) and their projection (straight red-lined segments) on a single dimension (curved red line), generated by training a single auto-associative artificial neural network on the data. To visualize this single dimension, expression profiles are shown in a three-dimensional (3D) space obtained through principal component analysis (PC1–3 represent the first three principal components). Expression profiles associated with one end of the single dimension are consistently associated with germination conditions (left circle), whereas those from the opposite end are associated with sporulation conditions (right circle). **b**, Single dimensions after training 100 independent auto-associative artificial neural networks on the data. The dark red line in **a** is randomly chosen from the 100 lines shown in **b** (see Methods).

subsequent germination, such as germinant receptors. Figure 4a shows that, in agreement with our expectations, the lifestyle effector genes are sequentially expressed in pseudo-time. First, cells express motility genes, then genes involved in colony formation and, finally, genes required for sporulation. Thus, our life cycle analysis, which is based on expression data alone, shows that the lifestyles of *B. subtilis* appear as clearly distinct and consecutive life stages.

We next examine the expression of upstream regulators during the *B. subtilis* life cycle. For this analysis, we consider all 40 regulators in the global transcription network that have 20 or more downstream regulatory targets, which includes the global regulators shown in Fig. 2. Following stringent significance criteria (see Methods), we identify 25 regulators that show significant expression changes in pseudo-time (Supplementary Figs. 12 and 13). Figure 4b,c shows the expression peaks and ranges for each of these regulators. Just like the effectors, the regulators allow us to distinguish the distinct life stages of *B. subtilis*. In brief, the motile life stage is associated with the expression of *sigD*, *swrAA* and *degU*[61]; the colony-forming life stage with the expression of *remA*, *comA*, *sigH*, *comK* and *spo0A*; and the sporulation life stage with the expression of *sigE*, *sigF*, *sigG*, *spoIID*, *sigK*, *gerE* and *spoVT*. Supplementary Figs. 14–19 show the expression profiles at six pseudo-time points highlighted in Fig. 4b, and they underscore that distinct regulatory modules are expressed serially in pseudo-time. The same life stages can also be identified in time series data of other Bacillales strains and species, although at much lower temporal resolution (Supplementary Fig. 20).

The order in which regulators are expressed over pseudo-time corresponds closely to what is known from studies that focus on the expression of individual regulators. For example, the expression patterns in Fig. 4b are consistent with the endosporulation programme (Supplementary Text 3): first, *spo0A* is expressed which leads to the expression of both *sigF* and *sigE*, which results in the expression of *sigG* and *sigK*. In addition, as observed in experimental studies[47], repressors are expressed early in the life cycle of *B. subtilis* (for example, AbrB and PurR) during the prosperous conditions that trigger germination. Finally, Fig. 4b shows that regulators that control life stage transitions, such as DegU and Spo0A, are expressed at the intersection between life stages[30,61–63]. Both these regulators induce different life stages depending on the level at which they are expressed and phosphorylated. DegU induces motility when phosphorylated at low levels and colony formation at high levels[30,61]. Spo0A induces colony formation when phosphorylated at low levels and sporulation at high levels[17,62,64,65].

**Mosaic evolution of life cycles in the Bacillales.** The modular organization of the *B. subtilis* life cycle, rooted in the serial expression of global regulators, raises the question: to what extent does life cycle modularity also affect evolution? To address this question, we study life cycle divergence in the Bacillales, the phylogenetic order to which *B. subtilis* belongs[66]. Specifically, we compare each of 384 complete genome sequences that cover the phylogenetic breadth of the Bacillales (Supplementary Fig. 21 and Supplementary Text 4) to the reference genome of *B. subtilis*[67,68]. For each pairwise comparison, we determine the presence of orthologous genes using bi-directional best BLAST hits[69–71] (Supplementary Text 4). To study the conservation of life stages, we focus on the life cycle regulators we identified in Fig. 4b, as well as on the downstream target genes they regulate. Specifically, we determine for each regulator how well the downstream regulon is conserved compared to the average conservation of genes across the entire genome. For our purpose, conservation is measured by the absence and presence of genes. We reason that weakly conserved regulons—with a high fraction of gene loss—may indicate the loss of a particular life stage whereas strongly conserved regulons—with a low fraction of gene loss—suggest its conservation.

Figure 5a shows the phylogenetic tree of the Bacillales together with the conservation of regulons plotted across the life cycle timeline. Weakly conserved regulons are shown in dark orange and strongly conserved regulons in dark blue. The phylogenetic data reveal a striking mosaic pattern, in which blocks of regulons are conserved and lost in a discrete fashion across species. These mosaic blocks correspond to the distinct life stages. In particular, the motility and sporulation life stages show highly repeatable and parallel losses across taxonomic families. Consistent with previous phylogenetic studies[66,69,72,73] (Supplementary Table 7), motility is

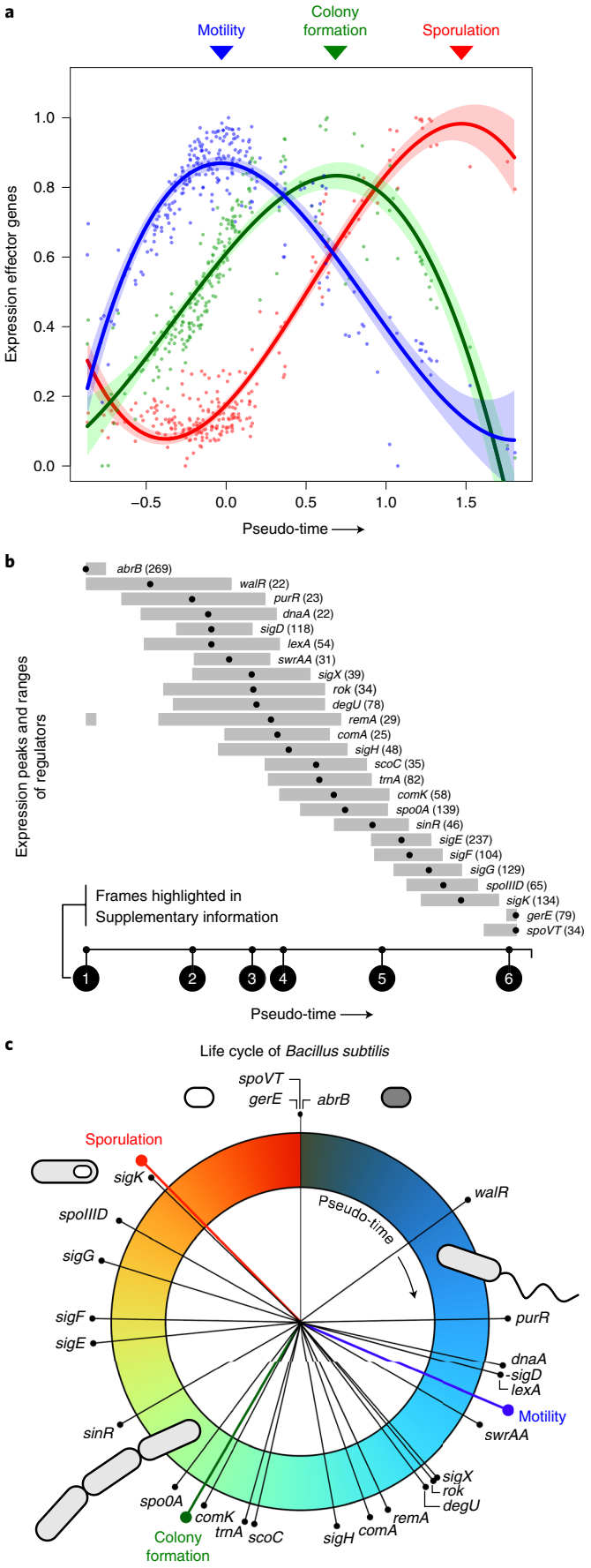


**Fig. 4 | Expression of life stages and regulators in pseudo-time. a**, Average expression level of several classes of genes in pseudo-time. Blue, motility genes; green, genes involved in colony formation and related starvation phenotypes; red, sporulation genes. Average expression levels are normalized between zero and one. Lines and shading indicate polynomial regression curves and 95% confidence intervals, respectively (blue, $R^2 = 0.75$, $P < 10^{-10}$; green, $R^2 = 0.73$, $P < 10^{-10}$; red, $R^2 = 0.92$, $P < 10^{-10}$). Upper triangles correspond to expression peaks. **b**, Expression of regulators in pseudo-time. For each regulator, a black circle denotes its expression peak and a grey bar the range of upper 5% expression levels, based on a polynomial regression (selection criteria, $R^2 > 0.3$ and $P < 10^{-5}$; see also Supplementary Figs. 12 and 13). Expression states in time frames 1–6 are highlighted in Supplementary Figs. 14–19. **c**, Schematic depiction of reconstructed life cycle: colours show life stages, obtained by converting the colour and expression levels in **a** into red/green/blue colour mapping; black lines show expression peaks of regulators (black dots in **b**), coloured lines show expression peaks of effector gene categories (triangles in **a**).

lost independently at least three times whereas sporulation is lost at least five times. In some species, only one of the life stages is lost (Fig. 5b; for example, *Exiguobacterium sibiricum*, *Listeria monocytogenes* and *Bacillus mycoides*), whereas in others both are lost (for example, *Staphylococcus aureus*). Importantly, the observed life stage losses cannot be attributed to a lack of phylogenetic signal since the life cycles of species both closely and distantly related to *B. subtilis* show very similar conservation patterns. Also, when corrected for phylogenetic distance, life stage conservation remains modular (Supplementary Fig. 27)—that is, regulons expressed in the same life stage have similar conservation patterns.

We hypothesized that the mosaic conservation pattern directly reflects the modular organization of the life cycle, in which life stages are conserved or lost depending on their utility in a species' habitat[74–78]. For example, some of the taxonomic families that lost sporulation have a symbiotic lifestyle (Staphylococcaceae and Listeriaceae). Their ancestors presumably underwent a lifestyle switch from a free-living planktonic lifestyle, in which sporulation is essential for dispersal, to a symbiotic lifestyle. In association with the host, sporulation was probably useless and therefore lost[66,79]. Laboratory evolution experiments show that *B. subtilis* can lose its capacity to sporulate when grown under an excess of nutrients for thousands of generations[80–83]. Although such conditions are highly unlikely to occur in nature, these experiments illustrate that life stages could in principle be lost through the accumulation of neutral mutations. Rather, we wanted to explore whether sporulation could also be lost under a selective regime that simply favours one of the other life stages, without the need to impose artificial laboratory conditions such as excessively long periods of nutrient excess. To this end, we designed an evolution experiment that favours colony growth, the life stage that is conserved among all Bacillales species, including those with a symbiotic lifestyle. We reasoned that, by selecting for colony growth, mutations that would promote growth by mitigating sporulation could readily fixate.

In total, we evolved eight Bacillales strains and species. These included *B. subtilis* and four of its close relatives (*B. subtilis*, *B. subtilis spizizenni*, *Bacillus licheniformis* and *Bacillus amyloliquefaciens*), as well as three more distantly related species (*Bacillus megaterium*, *Bacillus cereus* and *Bacillus thurigiensis*). For each strain or species, we grew cells on a nutrient-poor solid growth medium and, every week, transferred cells from the outermost colony edge to a fresh growth medium, thereby allowing the colony to evolve (Fig. 6a). In total, we transferred cells ten times, which resulted in 11 consecutive growth cycles. Each growth cycle is equivalent to approximately 20–25 generations. From analogous experiments in other species, we know that this is sufficient to observe a selective response[84–86].

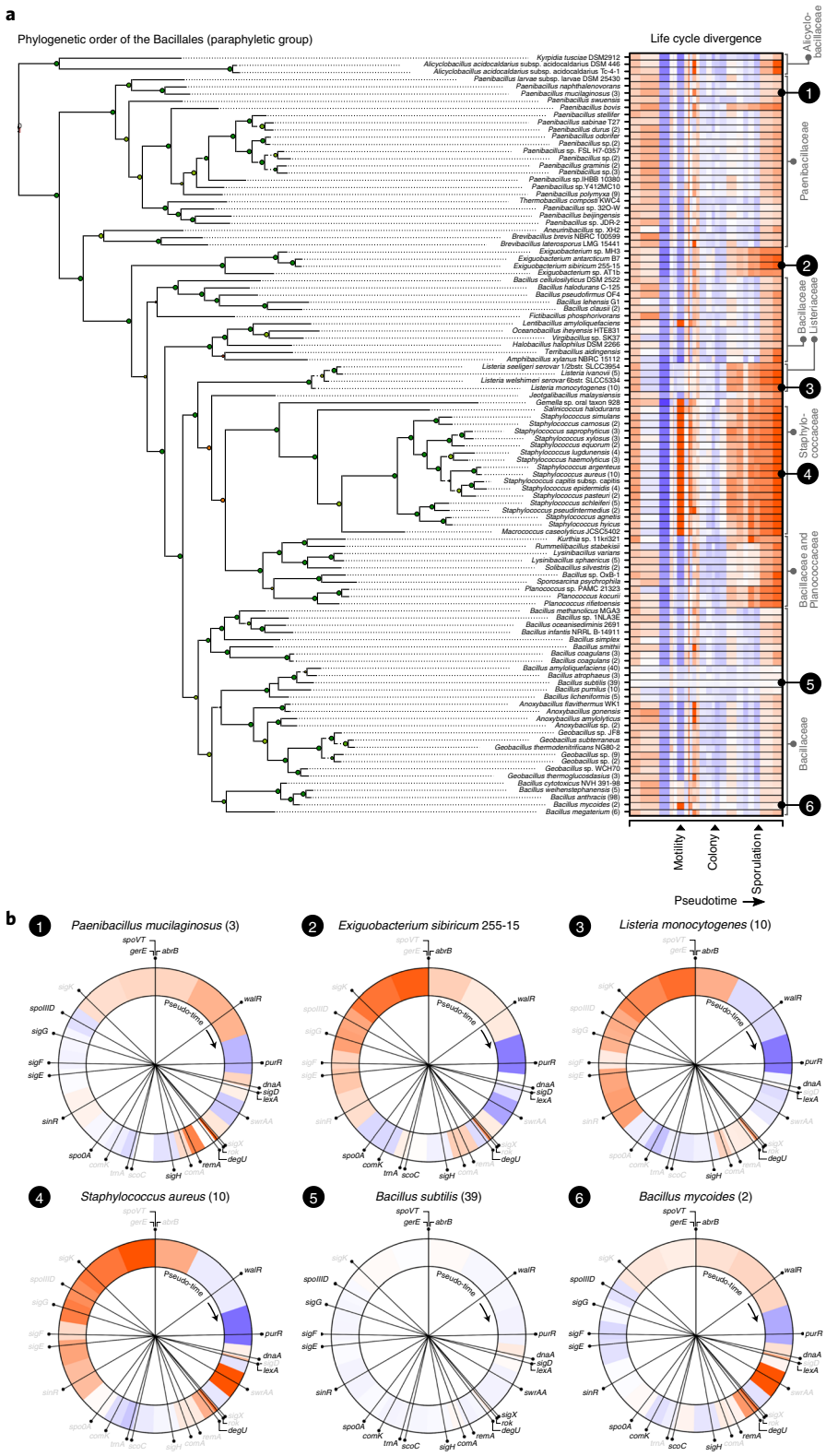


**Fig. 5 | Mosaic conservation patterns in the phylogenetic order of the Bacillales. a**, Phylogenetic tree of the Bacillales with conservation of life cycle regulons plotted in pseudo-time. Horizontal coloured bars to the right of each species indicate conservation of regulons corresponding to the regulators shown in Fig. 4b, relative to the average conservation of genes across the entire genome: dark orange, lower conservation of regulon than average conservation; dark blue, higher conservation of regulon than average conservation (white, average conservation). Some branches show the average conservation pattern for multiple, closely related strains (as indicated by the numbers in parentheses following species names). Sizes and colours of circles at phylogenetic branches show branch support: green dots indicate strongly supported branches whereas red dots indicate weakly supported branches (Supplementary Text 4). **b**, Example of life cycle divergence in six species (1–6), using the same colour scheme as in **a** but with the circular life cycle depiction from Fig. 4c. Numbers in parentheses next to each species name indicate the number of genomes analysed for that species. When the majority of genomes associated with a particular species lack an orthologous gene for a regulator, the name of that regulator is shown in grey in the corresponding life cycle.

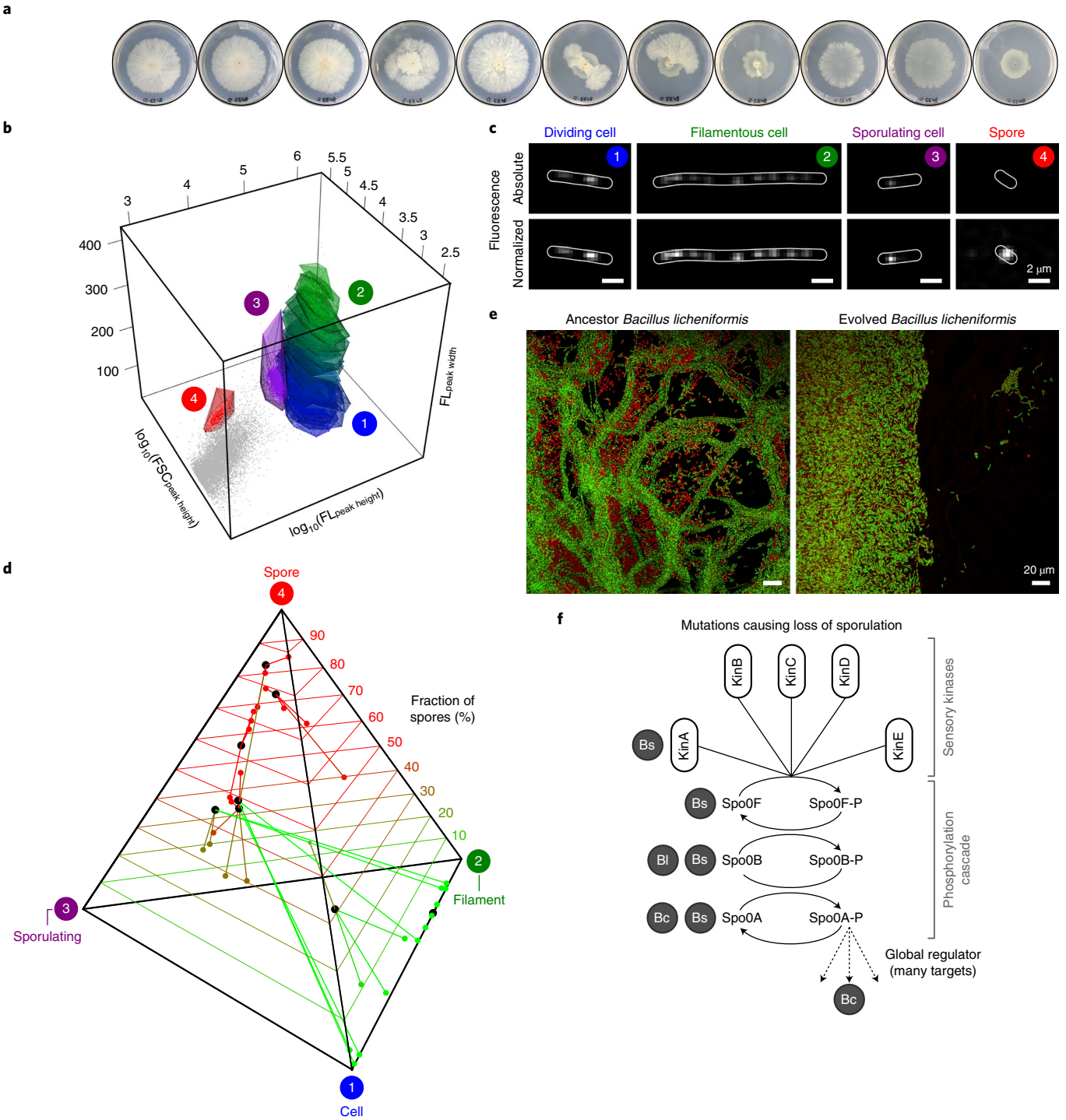


**Fig. 6 | Rapid mosaic adaptation in experiment evolution with Bacillales strains and species. a**, Example of colony growth in *B. licheniformis* on Petri dishes (pictured) during the evolution experiment. From left to right, colony growth in the 11 consecutive growth cycles. **b**, Flow cytometry results of cells obtained from *B. licheniformis* colony edge after 4 d of growth, plotted by forward scatter (FSC), height of fluorescent signal ($FL_{peak\ height}$) and width of fluorescent signal ($FL_{peak\ width}$). Flow cytometry events are gated in four groups, as shown by the 3D convex hulls (see Methods and Supplementary Fig. 28): (1) single cells, gated by low FSC, high $FL_{peak\ height}$ and low $FL_{peak\ width}$; (2) filamentous cells, gated by low FSC, high $FL_{peak\ height}$ and high $FL_{peak\ width}$; (3) sporulating cells, gated by high FSC, high $FL_{peak\ height}$ and low $FL_{peak\ width}$; and (4) spores, gated by high FSC, low $FL_{peak\ height}$ and low $FL_{peak\ width}$. Filaments are shown in separate sub-groups, from the shortest (blue) to the longest (green) filaments. Grey dots show debris and dead cells. **c**, Imaging cytometry pictures corresponding to the four cell states. Normalized fluorescence values show images with maximized contrast; white lines denote cell outlines. **d**, Tetrahedron showing evolutionary changes in cell state composition at colony edge. For simplicity, and because germinating spores were absent in most samples, spores and germinating spores are pooled. Ancestors are denoted by black dots and lines show change in cell state composition in evolved populations. Line colours denote the fraction of spores present in the evolved populations: red, 100%; green, 0%. **e**, Confocal scanning laser microscopy images showing loss of sporulation at the colony edge in the evolved *B. licheniformis* colonies: green, SYBR Green-stained DNA; red, MitoTracker Deep Red FM-stained cell membranes. Spores can be recognized by the presence of red membrane stain and absence of green DNA stain (Supplementary Fig. 33). **f**, Schematic overview of mutations in regulatory cascade that caused the loss of sporulation across species (see also Supplementary Fig. 3). Bs, *B. subtilis*; Bl, *B. licheniformis*; Bc, *B. cereus*.

Our selective regime favours advanced colony spreading, which may result from increased cell division rates, decreased lag times and altered growth patterns, such as filamentous growth. Because we evolved four replicates of each strain or species, our experiment comprised a total of 32 evolving populations.

We examined the ancestral and evolved populations using flow cytometry, which enabled us to examine the phenotypes of millions of individual cells by distinguishing five distinct cell states (see Methods): (1) single cells; (2) filamentous cells; (3) sporulating cells; (4) spores; and (5) germinating spores. Figure 6b,c shows an example of flow cytometry results from the colony edge of *B. licheniformis*, in which only the first four cell states are present (see also Supplementary Figs. 28 and 29). We compared the cell state composition of the ancestral and evolved populations at both the colony centre and edge, using at least three biological replicates per colony. This resulted in a total of 286 flow cytometry samples. The tetrahedron in Fig. 6d shows the evolutionary changes in cell state composition at the colony edge. Two species, *B. megaterium* and *B. amyloliquefaciens*, showed no evolutionary changes in either cell state composition or colony size. Of the remaining 24 evolved populations, 13 showed a significant reduction in sporulation whereas only one evolved a higher rate of sporulation (Supplementary Figs. 30 and 31). In most cases the reduction in sporulation was associated with larger colonies, which suggests that the reduction was adaptive for colony growth.

Strikingly, even though our evolution experiment lasted for only 11 weeks, sporulation was completely lost in five populations (Supplementary Fig. 32): one population of *B. licheniformis* (Fig. 6a,e), two populations of *B. subtilis* and two populations *B. cereus*. These losses often occurred within the first few growth cycles and were sometimes associated with strong filamentous growth, which is known to facilitate colony spreading[20] (Supplementary Fig. 33). To identify the genetic changes that underlie these remarkably rapid losses, we sequenced clones before and after the loss of sporulation for each of the five populations (see Methods). This revealed a striking genetic parallelism. In total, we identified seven mutations underlying the loss of sporulation (Supplementary Table 9). Six of these mutations affected the same global regulator, Spo0A, by preventing either its expression or activation (that is, by affecting the phosphorylation cascade; Fig. 6f). Spo0A plays a pivotal role in the life stage transition from colony growth to sporulation—without the activation of Spo0A, cells cannot sporulate (Fig. 4b and Supplementary Text 3). Thus, the evolution experiment corroborates our findings at the phylogenetic level and shows that selection in favour of colony growth is sufficient to drive the loss of sporulation. Moreover, the observed genetic parallelism again stresses the key importance of the life cycle, and its modular organization, in shaping evolutionary change[12,62].

## Discussion

We conclude that a life cycle perspective is critical for our understanding of both transcriptional regulation and evolution. The life cycle of *B. subtilis* is organized in a modular fashion, where the alternating life stages come about through the serial expression of global regulators and their respective regulatory modules[32,48]. We were able to reproduce the expression changes underlying life cycle progression by sorting a large number of previously acquired expression profiles along a single dimension, using machine learning. The simplicity of our methods suggests that cells also process and encode global regulatory information along one or few dimensions[87,88]. Indeed, global regulators often function as intermediates in regulatory cascades. That is, their activity and expression rely on many upstream sensory proteins and they affect many downstream effector proteins. As such, global regulators may reduce the dimensionality of the environment sensed by a cell onto a single dimension, which determines whether a downstream response is elicited[87–89].

The modular organization of the life cycle probably results from the alternating environmental conditions to which Bacillales species are exposed in nature[90–93]. Our experiments show that selection in favour of one life stage could quickly result in the loss of another. This suggests that life stages can be conserved only when species frequently encounter conditions that favour their expression[94–96]. Because life stages are generally well conserved across the Bacillales, which occupy a diverse range of habitats[97], they probably convey an advantage in many different environments. In other words, life stages do not express specialist phenotypes beneficial in a few environments only, but general-purpose phenotypes that are useful across a wide range of ecological habitats[98].

Our results also show that the modular organization of the life cycle has important implications for evolution. It provides common routes of adaptive change[99–103], which lead to parallelism at widely different time scales, from the few weeks of our evolution experiment to the millions of years of Bacillales species' divergence[66,69,73]. We postulate that life stages might typically be lost in the following way: first, species acquire mutations that inactivate a global regulator, thereby preventing the expression of a life stage. These mutations can fixate whenever a life stage becomes maladaptive, as in our evolution experiment. Second, through the accumulation of mutations, the regulatory modules that were initially controlled by the global regulator will gradually deteriorate, eventually resulting in mosaic conservation patterns as observed in our phylogenetic analysis[80,81].

Our study was necessarily restricted to the stages present in the life cycle of *B. subtilis* for which we had sufficient expression data for life cycle reconstruction. This does not imply that these are the only life stages present among the Bacillales. There exist Bacillales species with much larger genomes than that of *B. subtilis*[72,104]. These species often have many more sigma factors as well, sometimes even twice as many as *B. subtilis*[105,106] (Supplementary Fig. 34). Since sigma factors are known to play a key role in global transcriptional regulation[40–42] (Fig. 2), it is plausible that these Bacillales species express life stages that are absent in *B. subtilis*. At this moment, we can only speculate about the nature of such life stages. They might be involved in alternative types of colony growth or perhaps even primitive forms of fruiting body formation, a common bacterial adaptation to facilitate spore dispersal[5,107]. Exploring such putative life stages is an exciting task for future work.

## Methods

**Widefield images of *B. subtilis* in Fig. 1.** We acquired the images of *B. subtilis* NCBI 3610 in Fig. 1 from two different culturing conditions. We harvested motile cells from a liquid lysogeny broth culture (37 °C, 220 r.p.m.) using serial passaging at early exponential growth ($OD_{600}$ = 0.004), which results in a population of solely motile cells (for details on the passaging see ref. [20]). To acquire images of colony formation and sporulation, we grew cells on chemically defined agar medium at 37 °C, as described by in ref. [108], in which cells at the colony edge form filaments whereas starved cells in the colony centre sporulate. We used a Nikon Eclipse TE2000-U microscope equipped with a Nikon ×60/1.4 numerical aperture (NA) PlanApo oil objective to acquire phase contrast images.

**Global transcription network of *B. subtilis*.** We compared four independent reconstructions of the global transcription networks of *B. subtilis*: (1) the database of transcriptional regulation in *B. subtilis* (DBTBS) from ref. [28]; (2) the SubtiWiki database of transcriptional regulation[35]; and the transcriptional regulation network of ref. [34], both before (3) and after (4) inferring new regulatory interactions. For more details on these reconstructions, see Supplementary Text 1 and Supplementary Table 1. For each pair of regulators, we compared the observed fraction of coregulated genes to the expected fraction of coregulated genes. We calculated the expected overlap between regulons by randomizing the identity (but not the number) of genes under their regulatory control. In other words, when two regulators have large regulons the expected fraction of coregulated genes will increase. We determined the grouping of regulators based on the overlap between their respective regulons (Fig. 2 and Supplementary Fig. 2) through complete-linkage clustering of the ratio between the observed and expected overlap of regulons, using the stats library in R v.3.4.3.

**Life cycle reconstruction.** *Gene expression data and lifestyle genes.* To reconstruct the timeline of the *B. subtilis* life cycle, we used the processed gene expression profiles provided in ref. [42]. However, since those data were acquired from two different strains, we had to discard some gene expression profiles from our analysis. Supplementary Table 2 shows the list of profiles included in our analysis (see also Supplementary Table 1 in ref. [42]). Furthermore, we specifically wanted to focus on the expression of genes that are important for lifestyle transitions in *B. subtilis*. To determine these lifestyle genes, we made use of three information sources: (1) reviews that have extensively discussed the gene regulatory network underlying lifestyle transitions in *B. subtilis*[8,18,19,22,31,35,55–57]; (2) the SubtiWiki database, which provides a complete functional annotation of *B. subtilis* genes[35]; and (3) reconstructions of the global transcription networks of *B. subtilis*, as described in Methods and Supplementary Text 1. In our analysis we included only genes that were consistently associated with lifestyle biology in all three information sources (see Supplementary Table 3 and Supplementary Figs. 3 and 4 for an overview of these genes). However, we note that our results are not dependent on our specific selection of lifestyle genes because we obtained qualitatively similar results for different selections of such genes (Supplementary Fig. 7).

*Neural network analysis.* Before applying non-linear dimensionality reduction[58] with auto-associative artificial neural networks[59,60], we preprocessed the expression data by normalizing the expression of each gene between zero (minimum expression level that has been measured) and one (maximum expression level that has been measured). We then applied an orthogonal transformation to the data, in the form of a principal component analysis, and used the first 12 principal components (that is, linearly uncorrelated variables) to train the auto-associative artificial neural networks (applying a standard principal component analysis before training the artificial neural network strongly accelerates training). Each auto-associative artificial neural network has the same symmetrical topology consisting of five layers (Supplementary Fig. 5). The first two layers compress the gene expression data onto the central node in the third layer of the network, which represents the single dimension, while the subsequent two layers de-map the data from the central node into the output layer. The output layer represents gene expression levels as predicted by the trained network. We trained the neural network such that the input of the network (that is, observed gene expression levels) matches the output (that is, predicted gene expression levels) as closely as possible. To this end, we made use of the Nonlinear PCA toolbox in Matlab, as developed in refs. [59,60]. We trained the networks with a weight decay coefficient of 0.005 and a maximum of 1,000 iterative steps. At each iteration, gene expression profiles are first compressed onto the central node and subsequently expanded towards the output nodes. The network weights are then simultaneously updated using a backward propagation algorithm that minimizes the sum of squared errors between the network input and output[59]. Importantly, alternative weight decay coefficients and numbers of iterative steps yield qualitatively similar results (Supplementary Fig. 7). In total, we trained 100 artificial neural networks on the entire set of expression profiles. These profiles were sorted according to their average mapping on the resulting 100 single dimensions (Supplementary Fig. 6). Since the single dimension strongly correlates with time (Supplementary Figs. 10 and 11), we refer to this dimension as pseudo-time. Ten expression profiles could not robustly be mapped to the pseudo-time dimension, given the large standard deviation (>0.4) in their mapping on the 100 independently produced single dimensions (see Supplementary Fig. 6); we excluded these profiles from further downstream analyses. These corresponded to the following environmental conditions (see also Supplementary Table 2): confluent colonies on lysogeny broth plates (BC_1; BC/t_1; BC/t_2; BC_2), high osmolarity stress (HiOs_1; HiOs_2; HiOs_3) and stationary stage cultures from M9 medium (M9stat_1; M9stat_2; M9stat_3). Finally, we also explored an alternative type of dimensionality reduction by training auto-associative artificial neural networks that produce circular dimensions, as shown in ref. [109]. However, this alternative method did not yield robust results probably because there were insufficient data between the latest stages of sporulation and the earliest stages of germination.

*The expression of effector genes and regulators in pseudo-time.* We define effector genes as genes that are not involved in information transduction, but instead form the downstream target genes of the regulatory cascades underlying lifestyle transitions. As such, effector genes are responsible for expressing the phenotype of a cell (for example, flagella formation, chemotaxis, surfactin production, antimicrobial production, polysaccharide production, spore coat formation). We examined the expression of effector genes and regulators in pseudo-time. We focused specifically on effector genes underlying lifestyle transitions (Supplementary Fig. 4) and categorized these into three groups: (1) those underlying motility; (2) those underlying colony formation and related starvation phenotypes (for example, antimicrobial production, competence); and (3) those underlying sporulation. This categorization is based on the functional annotation of genes in the SubtiWiki database as well as of their upstream regulators (Supplementary Table 3). We normalized the expression levels of each gene between zero and one. Figure 4a gives equal weight to each gene by showing the average normalized expression of the genes assigned to a given lifestyle category over pseudo-time.

For our analysis of regulators, we focused on those present in the global transcription network of *B. subtilis* with 20 or more downstream regulatory targets, regardless of their function. For each of these regulators ($n = 40$), we normalized the expression level between zero and one and examined how this expression level changed over pseudo-time through a polynomial regression analysis. Specifically, we fitted first-, second- and third-order polynomial regressions through each regulator's expression trajectory and determined the best fit, based on the Bayesian information criterion. We used only those 25 regulators for which the polynomial regression met stringent significance criteria ($R^2 > 0.3$, $P < 10^{-5}$) for further downstream analyses. We determined the expression peak as well as the interval in pseudo-time at which each of these regulators reached the top 5% of their expression levels, as shown in Fig. 4b and Supplementary Figs. 12 and 13.

**Comparative genomics and phylogenetic analysis.** Supplementary Table 5 shows the list of genomes included in our comparative genomic analysis of the order Bacillales (NCBI taxonomy identifier 1385). We included all available genomes of the highest assembly level (NCBI assembly level = Complete Genome), except for the genomes of *S. aureus* and *L. monocytes*, from which we randomly selected only ten genomes each because otherwise they would have been highly over-represented. In addition, we excluded genomes from our analysis for which important information, such as annotated protein-coding sequences, was lacking. We compared all genomes to the same reference genome of *B. subtilis* 168 (AL009126 (refs. [67,68])). In line with previous studies[69–71], we used bi-directional-best BLAST hits as a proxy for gene orthology, with conservative cut-offs for protein BLAST hits ($<1 \times 10^{-10}$ and gene overlap (that is, qcovs) >70% (ref. [110]); the percentage of identical amino acids is not reported).

To construct the phylogenetic tree of our Bacillales species, we used a predefined set of 87 marker genes as identified in refs. [111,112]. Based on HMMER profiles (HMMER3.0 (refs. [113,114])), we identified these marker genes in each of our genomes (Supplementary Table 5). Since not all marker genes were universally present among our genomes, we reduced the list of marker genes using a few simple criteria, which are specified in Supplementary Text 4. For each marker gene, we aligned the sequences between genomes using MUSCLE v.3.8.31 (ref. [115]). Then, we trimmed the alignments using gBLOCKS v.0.91b to eliminate poorly aligned regions and end gaps[116]. Finally, we concatenated the marker gene sequences so that, for each genome, we ended up with a single concatenated sequence. We used these sequences to construct a maximum likelihood tree with RAxML v.8.2.10 (ref. [117]), based on the LG plus gamma model of evolution (PROTGAMMALG in the RAxML model section) and 200 bootstrap replicates.

To test the robustness of our phylogenetic tree construction we also systematically explored alternative methods, which are discussed in detail in Supplementary Text 4. In short, we varied the selection of marker genes using strict, intermediate or lenient criteria. We tried three types of alignment algorithms, MAFFT v.7.299, MUSCLE v.3.8.31 and PRANK[115,118–121]. And we tried two different trimming algorithms, trimAl v.1.2 and gBLOCKS v.0.91b[116,122,123]. The topologies of the phylogenetic trees that we constructed with these alternative methods were largely consistent with the tree shown in Fig. 5a and Supplementary Fig. 21 (Supplementary Text 4).

For visualization of the phylogenetic tree we employed the ETE 3.0 package[124,125], collapsing all branches with a phylogenetic distance <0.01 in the visualization. We performed all bioinformatic analyses in python, using the Biopython library[126], except for the phylogenetic principal component analysis (pPCA) shown in Supplementary Fig. 27, which we performed in R v.3.4.3 with the help of the libraries ape[127] and phytools[128]. For pPCA we assumed the Brownian motion evolutionary model.

**Evolution experiment.** We examined the evolution of eight Bacillales strains and species (Supplementary Table 8) by growing them for 11 consecutive growth cycles on a chemically defined growth medium in which we selected for advanced colony spread. The medium was derived from ref. [108] but specifically adapted to our needs, as specified in Supplementary Table 10. For every growth cycle we prepared our growth plates freshly, by solidifying our growth medium with 3% agar (No. A0949-500, AppliChem) and pouring it at a fixed temperature of 85 °C. We used 20 ml of growth medium for each standard Petri dish ($Ø = 94$ mm; No. 633181, Greiner Bio-One). After pouring we left plates on the bench for 60 min, sufficient for the medium to solidify, and then inverted them to prevent condensation droplets from falling onto the medium. We let the plates dry on the bench for approximately 20 h before inoculation of cells.

For each strain and species we prepared one starting population. We derived this starting population from an overnight culture, which was grown in 4 ml of our growth medium inoculated from a single clone. We archived the starting populations by mixing 500 µl of overnight culture with 500 µl 40% glycerol and freezing this in 2 ml CryoTubes (No. 72.380, Sarstedt) at −80 °C, having one starting population stock for each strain and species. We subsequently used the starting populations to initiate four independently evolving cell lines for each strain and species in the evolution experiment.

For the first growth cycle of the evolution experiment, we inoculated cells from the eight frozen starting populations on eight growth plates that we incubated

overnight at 37 °C. The next morning, we harvested cells from the plates with a 10-µl inoculation loop (No. 86.1562.010, Sarstedt) and resuspended them in 500 µl PBS. We subsequently normalized the cell density of the eight cell suspensions to an optical density ($\lambda = 600$ nm) of 0.01 ($OD_{600}$). From each of these normalized cell suspensions we inoculated four growth plates by pipetting 2 µl in the centre of the plate. This procedure ensures that the replicate populations of each strain and species are derived from the same clonal ancestral population. In total we inoculated 32 plates, four replicate cell lines for each strain and species. When the inoculation droplets had dried, we wrapped plates in parafilm and subsequently incubated them upside down for a week at 37 °C.

For the remaining ten growth cycles, we transferred cells every week using a standard procedure. With a 10-µl inoculation loop, we scraped cells either from the outermost colony edge or from an arbitrary edge if the colony was round. We resuspended these cells in 500 µl PBS buffer, normalized cell density to $OD_{600} = 0.01$ and plated 2 µl (corresponding to approximately $10^4$ cells) of cell suspension to start the next growth cycle. As in the first growth cycle, we waited until the inoculum droplets had dried then wrapped the plates in parafilm and incubated them at 37 °C. We used part of the remaining cell suspensions (200–300 µl) to archive the 32 cell lines by mixing them with 200 µl 40% glycerol in CryoTubes and storing these at −80 °C.

In principle the colony morphology alone is sufficient to discriminate between all strains and species, yet to completely exclude cross-contamination we also performed regular PCR checks on all 32 evolved cell lines (using GoTaq G2 Green Master Mix, No. M7822, Promega). We specifically amplified genes unique to each given strain or species (see Supplementary Table 11 for primer sets). In this way, we could discriminate between all strains and species, except for two strains that were too closely related. We checked each cell line about once per month and found no incidences of cross-contamination.

**Phenotyping.** *Flow cytometry and imaging cytometry.* For flow cytometry and imaging cytometry, we first grew colonies according to our standard growth cycle as described above. We harvested cells from either the colony edge (Fig. 6b and Supplementary Figs. 28, 29 and 31) or centre (Supplementary Figs. 29 and 32) using a 10-µl inoculation loop, then resuspended them in 500 µl PBS. For flow cytometry, we diluted cell suspensions to $OD_{600} = 0.001$ and, for imaging cytometry, we diluted them to $OD_{600} = 0.05$. To each resulting OD-normalized cell culture of 500 µl we added 5 µl of 100× SYBR Green I staining solution, consisting of SYBR Green I stock solution (No. S7563, Invitrogen, Thermo Fisher Scientific) dissolved in 10 mM Tris (pH 8.0) buffer, as described in the protocol of Thermo Fisher Scientific. We subsequently vortexed and incubated the samples either for 15 min at 37 °C on a heat block or for 20 min at 37 °C in an incubator. For flow cytometry, we used the BD Accuri C6 flow cytometer (Becton Dickinson, device used from Drinking Water Microbiology group at Eawag) to analyse 50 µl of each sample at the maximum flow rate of 66 µl $min^{-1}$. Between all samples we flushed the cytometer with 150 µl of filter-sterilized water (0.1 µm pore size, No. SLVV033RS, Millex-VV, Merck Millipore) or 70% ethanol solution. For imaging flow cytometry, we used the Image Stream X Mark II imaging flow cytometer (Amdis, Merck Millipore) with an extended-depth of field camera (provided by Cytometry Facility, University of Zürich). We analysed image cytometry data using IDEAS Software (Merck Millipore) and the Image Processing Toolbox of Matlab 2016 (MathWorks).

We analysed all flow cytometry data with R v.3.4.3 using the libraries SDMTools, alphahull and rgl[129–131]. Since cell sizes, spore sizes and genome contents differed slightly between strains and species, we had to optimize the gating procedure for each strain and species separately. Nevertheless, within each strain and species we applied the same gating on all samples. Generally, we applied the gating procedure in three steps. First, we discriminated between low- and high-fluorescent events. High-fluorescent events correspond to individuals for which the DNA could be stained, such as cells, filaments and sporulating cells. Low-fluorescent events correspond to individuals for which the DNA could not be stained, such as (germinating) spores (that is, SYBR Green staining cannot penetrate the spore coat). Second, we discriminated between samples with low and high forward scatter. The increased refraction index of dehydrated (fore-)spores led to higher forward scatter, which allowed us to gate for spores and sporulating cells[132]. For individuals with low forward scatter and high green fluorescence—that is, non-sporulating cells—we used the fluorescent width to establish filament size. The width shows the length of the filament through the number of stained genomes and thus cells. For all flow cytometry experiments, we calibrated the results with 30-µm fluorescent beads (BD CT&T RUO Beads, No. 661414, Becton Dickinson). We examined the accuracy of our gating using negative control samples (in which cells were filtered out), microscopy (for example, Fig. 6c and Supplementary Fig. 33) and image flow cytometry (for example, Fig. 6c). Note that, in Fig. 6b and Supplementary Fig. 28, we show the outcome of the gating procedure using 3D convex hulls that surround the data points clustered together after gating. All flow cytometry data are publically available through the FlowRepository[133,134] (No. FR-FCM-ZYVN).

*Camera.* For the entire duration of the evolution experiment, images were taken 6 d after inoculation (that is, 1 d before re-inoculation of cells onto fresh growth plates) using an EOS 1200D Canon Camera with EF-S 18–55-mm objective.

*Microscopy.* For microscopy, we cut the colony edge from the plate with a scalpel and placed it on a microscopy glass slide (Menzel Gläser 26 × 76 $mm^2$, No. 10144633, Thermo Fisher Scientific), after which we applied 50–100 µl of staining solution. This solution consists of PBS buffer with 2% v/v formaldehyde (No. 47608-1L-F, Sigma) and the stains 1× SYBR Green and 1× MitoTracker Deep Red FM (No. M22426, Thermo Fisher Scientific). We made fresh staining solution for every round of microscopy, from stock solutions of both stains that were prepared in accordance with the protocol of Thermo Fisher Scientific. Although colonies can be strongly hydrophobic towards the colony centre, at the edge the staining solution was readily absorbed. We used SYBR Green to stain the DNA (like in the case of flow cytometry) and MitoTracker Deep Red FM to stain the cell membrane and spore coat. After staining, we placed the glass slide in a sterile Petri dish and incubated this at 37 °C for 25 min. We subsequently applied one or two droplets of ProLong Diamond Antifade Mountant buffer (No. P36965, Thermo Fisher Scientific) and placed a cover slip on top (Menzel Microscope Coverslips 24 × 60 $mm^2$, No. 11778691, Thermo Fisher Scientific). For the curing process of the mountant buffer, we left the microscopy sample in darkness at room temperature for a period of 3 d. After this, we sealed the microscopy sample airtight using VALAP (a 1/1/1 mixture of lanolin, paraffin and petroleum jelly) and subsequently examined the colony edge with confocal scanning laser microscopy. Specifically, we used the CSLM SP8 (Leica) upright microscope at The Center for Microscopy and Image Analysis of the University of Zürich, with a Leica ×63/1.4 NA HCX PL APO CS2 oil objective, 8,000-Hz resonant scanner, z-Galvo stage insert (Leica) and solid-state diode lasers at 488 nm (20 mW) and 638 nm (30 mW).

**Whole-genomic sequencing.** In the evolution experiment, five populations completely lost sporulation—replicate 2 of *B. licheniformis*, replicates 1 and 4 of *B. subtilis* subspecies *subtilis* and replicates 1 and 2 of *B. cereus*. For each population, we selected four clones for whole-genome sequencing (Supplementary Table 9)—one clone from the ancestor, one from the population before the loss of sporulation and two from the population after the loss of sporulation. In total, we sequenced the genomes of 18 clones (3 ancestral clones + 3 × 5 derived clones). We selected clones from different growth cycles, depending on the growth cycle at which sporulation was lost, as verified by flow cytometry. For each clone, we extracted DNA using the Promega Wizard Genomic DNA Purification Kit (No. A1120) following the manufacturer's instructions, with the exception for the initial Gram-positive cell lysis step where we resuspended cell pellets in 480 µl of 50 mM EDTA (Ambion, No. AM9260G) and 120 µl 100 mg $ml^{-1}$ lysoszyme (Sigma, No. L4919) and incubated them for 60 min at 37 °C. We checked DNA quality using Qubit dsDNA BR Assay (Invitrogen, No. Q32853), a NanoDrop 1000 UV-VIS Spectrophotometer and a 0.7% agarose TBE gel. Subsequent library construction and whole-genome sequencing were performed by MicrobesNG (University of Birmingham, UK, https://microbesng.uk/), using Illumina HiSeq 2500 sequencing with 2 × 250-base pair paired-end reads. We used *breseq* v.0.32.0 (refs. [135,136]) to map the sequencing reads against the respective reference genomes (see Supplementary Table 9) and, subsequently, compared genomic differences among clones.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The flow cytometry data are publically available through the FlowRepository, accession No. FR-FCM-ZYVN. The sequencing reads are publically available on the European Nucleotide Archive (ENA) database, accession No. PRJEB32792. The remaining data are included in the Supplementary information or available through public repositories as mentioned in the Supplementary information.

Received: 3 October 2018; Accepted: 3 June 2019;
Published online: 22 July 2019

## References

1. Kolenbrander, P. E., Palmer, R. J., Periasamy, S. & Jakubovics, N. S. Oral multispecies biofilm development and the key role of cell–cell distance. *Nat. Rev. Microbiol.* **8**, 471–480 (2010).
2. Flores, E. & Herrero, A. Compartmentalized function through cell differentiation in filamentous cyanobacteria. *Nat. Rev. Microbiol.* **8**, 39–50 (2010).
3. McDougald, D., Rice, S. A., Barraud, N., Steinberg, P. D. & Kjelleberg, S. Should we stay or should we go: Mechanisms and ecological consequences for biofilm dispersal. *Nat. Rev. Microbiol.* **10**, 39–50 (2012).
4. Boutte, C. C. & Crosson, S. Bacterial lifestyle shapes stringent response activation. *Trends Microbiol.* **21**, 174–180 (2013).
5. Claessen, D., Rozen, D. E., Kuipers, O. P., Søgaard-Andersen, L. & van Wezel, G. P. Bacterial solutions to multicellularity: a tale of biofilms, filaments and fruiting bodies. *Nat. Rev. Microbiol.* **12**, 115–124 (2014).
6. van Gestel, J., Vlamakis, H. & Kolter, R. Division of labor in biofilms: the ecology of cell differentiation. *Microbiol. Spectr.* **3**, MB-0002–MB-2014 (2015).

7. Yan, J., Nadell, C. D. & Bassler, B. L. Environmental fluctuation governs selection for plasticity in biofilm production. *ISME J.* **11**, 1569–1577 (2017).
8. Mhatre, E., Monterrosa, R. G. & Kovács, Á. T. From environmental signals to regulators: modulation of biofilm development in Gram-positive bacteria. *J. Basic Microbiol.* **54**, 616–632 (2014).
9. Winslow, C. E. A. What do we mean by a bacterial life cycle? *Science* **81**, 314–315 (1935).
10. O'Toole, G., Kaplan, H. B. & Kolter, R. Biofilm formation as microbial development. *Annu. Rev. Microbiol.* **54**, 49–79 (2000).
11. Hammerschmidt, K., Rose, C. J., Kerr, B. & Rainey, P. B. Life cycles, fitness decoupling and the evolution of multicellularity. *Nature* **515**, 75–79 (2014).
12. Stragier, P. & Losick, R. Molecular genetics of sporulation in *Bacillus subtilis*. *Annu. Rev. Genet.* **30**, 297–341 (1996).
13. Curtis, P. D. & Brun, Y. V. Getting in the loop: regulation of development in *Caulobacter crescentus*. *Microbiol. Mol. Biol. Rev.* **74**, 13–41 (2010).
14. Norman, T. M., Lord, N. D., Paulsson, J. & Losick, R. Memory and modularity in cell-fate decision making. *Nature* **503**, 481–486 (2013).
15. Russell, J. R., Cabeen, M. T., Wiggins, P. A., Paulsson, J. & Losick, R. Noise in a phosphorelay drives stochastic entry into sporulation in *Bacillus subtilis*. *EMBO J.* **36**, 2856–2869 (2017). e201796988.
16. Kearns, D. B. A field guide to bacterial swarming motility. *Nat. Rev. Microbiol.* **8**, 634–644 (2010).
17. Grau, R. R. et al. A duo of potassium-responsive histidine kinases govern the multicellular destiny of *Bacillus subtilis*. *mBio* **6**, e00581 (2015).
18. Cairns, L. S., Hobley, L. & Stanley-Wall, N. R. Biofilm formation by *Bacillus subtilis*: new insights into regulatory strategies and assembly mechanisms. *Mol. Microbiol.* **93**, 587–598 (2014).
19. Mielich-Süss, B. & Lopez, D. Molecular mechanisms involved in *Bacillus subtilis* biofilm formation. *Environ. Microbiol.* **17**, 555–565 (2015).
20. van Gestel, J., Vlamakis, H. & Kolter, R. From cell differentiation to cell collectives: *Bacillus subtilis* uses division of labor to migrate. *PLoS Biol.* **13**, e1002141 (2015).
21. Beauregard, P. B., Chai, Y., Vlamakis, H., Losick, R. & Kolter, R. *Bacillus subtilis* biofilm induction by plant polysaccharides. *Proc. Natl Acad. Sci. USA* **110**, E1621–E1630 (2013).
22. Higgins, D. & Dworkin, J. Recent progress in *Bacillus subtilis* sporulation. *FEMS Microbiol. Rev.* **36**, 131–148 (2012).
23. Setlow, P. Spore germination. *Curr. Opin. Microbiol.* **6**, 550–556 (2003).
24. Smits, W. K., Kuipers, O. P. & Veening, J. W. Phenotypic variation in bacteria: the role of feedback regulation. *Nat. Rev. Microbiol.* **4**, 259–271 (2006).
25. Bonneau, R. et al. The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets *de novo*. *Genome Biol.* **7**, R36 (2006).
26. Aguilar, C., Vlamakis, H., Losick, R. & Kolter, R. Thinking about *Bacillus subtilis* as a multicellular organism. *Curr. Opin. Microbiol.* **10**, 638–643 (2007).
27. Veening, J. W., Smits, W. K. & Kuipers, O. P. Bistability, epigenetics, and bet-hedging in bacteria. *Annu. Rev. Microbiol.* **62**, 193–210 (2008).
28. Sierro, N., Makita, Y., de Hoon, M. & Nakai, K. DBTBS: a database of transcriptional regulation in *Bacillus subtilis* containing upstream intergenic conservation information. *Nucleic Acids Res.* **36**, D93–D96 (2008).
29. Fadda, A. et al. Inferring the transcriptional network of *Bacillus subtilis*. *Mol. Biosyst.* **5**, 1840–1852 (2009).
30. Kobayashi, K. Gradual activation of the response regulator DegU controls serial expression of genes for flagellum formation and biofilm formation in *Bacillus subtilis*. *Mol. Microbiol.* **66**, 395–409 (2007).
31. Lopez, D. & Kolter, R. Extracellular signals that define distinct and coexisting cell fates in *Bacillus subtilis*. *FEMS Microbiol. Rev.* **34**, 134–149 (2010).
32. Freyre-González, J. A. et al. Lessons from the modular organization of the transcriptional regulatory network of *Bacillus subtilis*. *BMC Syst. Biol.* **7**, 127 (2013).
33. Leyn, S. A. et al. Genomic reconstruction of the transcriptional regulatory network in *Bacillus subtilis*. *J. Bacteriol.* **195**, 2463–2473 (2013).
34. Arrieta-Ortiz, M. L. et al. An experimentally supported model of the *Bacillus subtilis* global transcriptional regulatory network. *Mol. Syst. Biol.* **11**, 839 (2015).
35. Michna, R. H., Zhu, B., Mäder, U. & Stülke, J. *Subti*Wiki 2.0—an integrated database for the model organism *Bacillus subtilis*. *Nucleic Acids Res.* **44**, D654–D662 (2016).
36. Lee, T. I. et al. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298**, 799–804 (2002).
37. Sorrells, T. R. & Johnson, A. D. Making sense of transcription networks. *Cell* **161**, 714–723 (2015).
38. Fang, X. et al. Global transcriptional regulatory network for *Escherichia coli* robustly connects gene expression to transcription factor activities. *Proc. Natl Acad. Sci. USA* **114**, 10286–10291 (2017).
39. Santos-Zavaleta, A. et al. RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in *E. coli* K-12. *Nucleic Acids Res.* **47**, D212–D220 (2019).
40. Haldenwang, W. G. The sigma factors of *Bacillus subtilis*. *Microbiol. Rev.* **59**, 1–30 (1995).
41. Gruber, T. M. & Gross, C. A. Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu. Rev. Microbiol.* **57**, 441–466 (2003).
42. Nicolas, P. et al. Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*. *Science* **335**, 1103–1106 (2012).
43. Feklístov, A., Sharon, B. D., Darst, S. A. & Gross, C. A. Bacterial sigma factors: a historical, structural, and genomic perspective. *Annu. Rev. Microbiol.* **68**, 357–376 (2014).
44. Veening, J. W. et al. Bet-hedging and epigenetic inheritance in bacterial cell development. *Proc. Natl Acad. Sci. USA* **105**, 4393–4398 (2008).
45. Marlow, V. L. et al. The prevalence and origin of exoprotease-producing cells in the *Bacillus subtilis* biofilm. *Microbiology* **160**, 56–66 (2014).
46. Vilain, S., Luo, Y., Hildreth, M. B. & Brozel, V. S. Analysis of the life cycle of the soil saprophyte *Bacillus cereus* in liquid soil extract and in soil. *Appl. Environ. Microbiol.* **72**, 4970–4977 (2006).
47. Otto, A. et al. Systems-wide temporal proteomic profiling in glucose-starved *Bacillus subtilis*. *Nat. Commun.* **1**, 137 (2010).
48. Omony, J., de Jong, A., Krawczyk, A. O., Eijlander, R. T. & Kuipers, O. P. Dynamic sporulation gene co-expression networks for *Bacillus subtilis* 168 and the food-borne isolate *Bacillus amyloliquefaciens*: a transcriptomic model. *Microb. Genom.* **4**, 1–13 (2018).
49. Bendall, S. C. et al. Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell* **157**, 714–725 (2014).
50. Anavy, L. et al. BLIND ordering of large-scale transcriptomic developmental timecourses. *Development* **141**, 1161–1166 (2014).
51. Haghverdi, L., Büttner, M., Wolf, F. A., Buettner, F. & Theis, F. J. Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* **13**, 845–848 (2016).
52. Levin, M. et al. The mid-developmental transition and the evolution of animal body plans. *Nature* **531**, 637–641 (2016).
53. Setty, M. et al. Wishbone identifies bifurcating developmental trajectories from single-cell data. *Nat. Biotechnol.* **34**, 637–645 (2016).
54. Marioni, J. C. & Arendt, D. How single-cell genomics is changing evolutionary and developmental biology. *Annu. Rev. Cell Dev. Biol.* **33**, 537–553 (2017).
55. Garrity, L. F. & Ordal, G. W. Chemotaxis in *Bacillus subtilis*: how bacteria monitor environmental signals. *Pharmacol. Ther.* **68**, 87–104 (1995).
56. McKenney, P. T., Driks, A. & Eichenberger, P. The *Bacillus subtilis* endospore: assembly and functions of the multilayered coat. *Nat. Rev. Microbiol.* **11**, 33–44 (2013).
57. Vlamakis, H., Chai, Y., Beauregard, P., Losick, R. & Kolter, R. Sticking together: building a biofilm the *Bacillus subtilis* way. *Nat. Rev. Microbiol.* **11**, 157–168 (2013).
58. Kramer, M. A. Nonlinear principal component analysis using autoassociative neural networks. *AIChE J.* **37**, 233–243 (1991).
59. Scholz, M., Fraunholz, M. & Selbig, J. in *Principal Manifolds for Data Visualization and Dimension Reduction* 44–67 (Springer, 2008).
60. Scholz, M. Validation of nonlinear PCA. *Neural Process. Lett.* **36**, 21–30 (2012).
61. Verhamme, D. T., Kiley, T. B. & Stanley-Wall, N. R. DegU co-ordinates multicellular behaviour exhibited by *Bacillus subtilis*. *Mol. Microbiol.* **65**, 554–568 (2007).
62. Fujita, M., González-Pastor, J. E. & Losick, R. High- and low-threshold genes in the Spo0A regulon of *Bacillus subtilis*. *J. Bacteriol.* **187**, 1357–1368 (2005).
63. Verhamme, D. T., Murray, E. J. & Stanley-Wall, N. R. DegU and Spo0A jointly control transcription of two loci required for complex colony development by *Bacillus subtilis*. *J. Bacteriol.* **191**, 100–108 (2009).
64. Branda, S. S., González-Pastor, J. E., Ben-Yehuda, S., Losick, R. & Kolter, R. Fruiting body formation by *Bacillus subtilis*. *Proc. Natl Acad. Sci. USA* **98**, 11621–11626 (2001).
65. Branda, S. S., Chu, F., Kearns, D. B., Losick, R. & Kolter, R. A major protein component of the *Bacillus subtilis* biofilm matrix. *Mol. Microbiol.* **59**, 1229–1238 (2006).
66. Galperin, M. Y. Genome diversity of spore-forming firmicutes. *Microbiol. Spectr.* **1**, 1–15 (2013).
67. Kunst, F. et al. The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature* **390**, 249–256 (1997).
68. Barbe, V. et al. From a consortium sequence to a unified sequence: the *Bacillus subtilis* 168 reference genome a decade later. *Microbiology* **155**, 1758–1775 (2009).
69. de Hoon, M. J. L., Eichenberger, P. & Vitkup, D. Hierarchical evolution of the bacterial sporulation network. *Curr. Biol.* **20**, R735–R745 (2010).

70. Wolf, Y. I. & Koonin, E. V. A tight link between orthologs and bidirectional best hits in bacterial and archaeal genomes. *Genome Biol. Evol.* **4**, 1286–1294 (2012).
71. Gabaldón, T. & Koonin, E. V. Functional and evolutionary implications of gene orthology. *Nat. Rev. Genet.* **14**, 360–366 (2013).
72. Galperin, M. Y. et al. Genomic determinants of sporulation in *Bacilli* and *Clostridia*: towards the minimal set of sporulation-specific genes. *Environ. Microbiol.* **14**, 2870–2890 (2012).
73. Abecasis, A. B. et al. A genomic signature and the identification of new sporulation genes. *J. Bacteriol.* **195**, 2101–2115 (2013).
74. Moran, N. A. Microbial minimalism: genome reduction in bacterial pathogens. *Cell* **108**, 583–586 (2002).
75. Makarova, K. et al. Comparative genomics of the lactic acid bacteria. *Proc. Natl Acad. Sci. USA* **103**, 15611–15616 (2006).
76. Wolf, Y. I. & Koonin, E. V. Genome reduction as the dominant mode of evolution. *BioEssays* **35**, 829–837 (2013).
77. Albalat, R. & Cañestro, C. Evolution by gene loss. *Nat. Rev. Genet.* **17**, 379–391 (2016).
78. Duar, R. M. et al. Lifestyles in transition: evolution and natural history of the genus *Lactobacillus*. *FEMS Microbiol. Rev.* **41**, S27–S48 (2017).
79. Sokurenko, E. V., Hasty, D. L. & Dykhuizen, D. E. Pathoadaptive mutations: gene loss and variation in bacterial pathogens. *Trends Microbiol.* **7**, 191–195 (1999).
80. Maughan, H. et al. The population genetics of phenotypic deterioration in experimental populations of *Bacillus subtilis*. *Evol. Int. J. Org. Evol.* **60**, 686–695 (2006).
81. Maughan, H., Masel, J., Birky, C. W. & Nicholson, W. L. The roles of mutation accumulation and selection in loss of sporulation in experimental populations of *Bacillus subtilis*. *Genetics* **177**, 937–948 (2007).
82. Maughan, H., Birky, C. W. & Nicholson, W. L. Transcriptome divergence and the loss of plasticity in *Bacillus subtilis* after 6,000 generations of evolution under relaxed selection for sporulation. *J. Bacteriol.* **191**, 428–433 (2009).
83. Brown, C. T. et al. Whole-genome sequencing and phenotypic analysis of *Bacillus subtilis* mutants following evolution under conditions of relaxed selection for sporulation. *Appl. Environ. Microbiol.* **77**, 6867–6877 (2011).
84. Velicer, G. J. & Yu, Y. T. N. Evolution of novel cooperative swarming in the bacterium *Myxococcus xanthus*. *Nature* **425**, 75–78 (2003).
85. van Ditmarsch, D. et al. Convergent evolution of hyperswarming leads to impaired biofilm formation in pathogenic bacteria. *Cell Rep.* **4**, 697–708 (2013).
86. Song, C., Kidarsa, T. A., van de Mortel, J. E., Loper, J. E. & Raaijmakers, J. M. Living on the edge: emergence of spontaneous *gac* mutations in *Pseudomonas protegens* during swarming motility. *Environ. Microbiol.* **18**, 3453–3465 (2016).
87. Friedlander, T., Mayo, A. E., Tlusty, T. & Alon, U. Evolution of bow-tie architectures in biology. *PLoS Comput. Biol.* **11**, e1004055 (2015).
88. Yan, J. et al. Bow-tie signaling in c-di-GMP: machine learning in a simple biochemical network. *PLoS Comput. Biol.* **13**, e1005677 (2017).
89. Kitano, H. Biological robustness. *Nat. Rev. Genet.* **5**, 826–837 (2004).
90. Bonner, J. T. *The Evolution of Complexity by Means of Natural Selection* (Princeton Univ. Press, 1988).
91. Kashtan, N. & Alon, U. Spontaneous evolution of modularity and network motifs. *Proc. Natl Acad. Sci. USA* **102**, 13773–13778 (2005).
92. Wagner, G. P., Pavlicev, M. & Cheverud, J. M. The road to modularity. *Nat. Rev. Genet.* **8**, 921–931 (2007).
93. Espinosa-Soto, C. & Wagner, A. Specialization can drive the evolution of modularity. *PLoS Comput. Biol.* **6**, e1000719 (2010).
94. Lande, R. Evolution of phenotypic plasticity and environmental tolerance of a labile quantitative character in a fluctuating environment. *J. Evol. Biol.* **27**, 866–875 (2014).
95. Huang, B. & Mackem, S. Evolutionary developmental biology: use it or lose it. *Nature* **511**, 34–35 (2014).
96. Siljestam, M. & Östman, Ö. The combined effects of temporal autocorrelation and the costs of plasticity on the evolution of plasticity. *J. Evol. Biol.* **30**, 1361–1371 (2017).
97. Mandic-Mulec, I., Stefanic, P. & van Elsas, J. D. Ecology of Bacillaceae. *Microbiol. Spectr.* **3**, TBS-0017–TBS-2013 (2015).
98. Parter, M., Kashtan, N. & Alon, U. Facilitated variation: how evolution learns from past environments to generalize to new environments. *PLoS Comput. Biol.* **4**, e1000206 (2008).
99. Riedl, R. A systems-analytical approach to macro-evolutionary phenomena. *Q. Rev. Biol.* **52**, 351–370 (1977).
100. Wagner, G. P. & Altenberg, L. Perspective: complex adaptations and the evolution of evolvability. *Evolution* **50**, 967–976 (1996).
101. Schluter, D. Adaptive radiation along genetic lines of least resistance. *Evolution* **50**, 1766–1774 (1996).
102. Watson, R. A. & Szathmáry, E. How can evolution learn? *Trends Ecol. Evol.* **31**, 147–157 (2016).
103. Uller, T., Moczek, A. P., Watson, R. A., Brakefield, P. M. & Laland, K. N. Developmental bias and evolution: a regulatory network perspective. *Genetics* **209**, 949–966 (2018).
104. Mead, D. A. et al. Complete genome sequence of *Paenibacillus* strain Y4.12MC10, a novel *Paenibacillus lautus* strain isolated from Obsidian hot spring in Yellowstone National Park. *Stand. Genom. Sci.* **6**, 381–400 (2012).
105. van Nimwegen, E. in *Power Laws, Scale-Free Networks and Genome Biology* 236–253 (Springer, 2006).
106. Cordero, O. X. & Hogeweg, P. Large changes in regulome size herald the main prokaryotic lineages. *Trends Genet.* **23**, 488–493 (2007).
107. Barka, E. A. et al. Taxonomy, physiology, and natural products of *Actinobacteria*. *Microbiol. Mol. Biol. Rev.* **80**, 1–43 (2016).
108. Fall, R., Kearns, D. B. & Nguyen, T. A defined medium to investigate sliding motility in a *Bacillus subtilis* flagella-less mutant. *BMC Microbiol.* **6**, 31 (2006).
109. Scholz, M. & Fraunholz, M. J. A computational model of gene expression reveals early transcriptional events at the subtelomeric regions of the malaria parasite, *Plasmodium falciparum*. *Genome Biol.* **9**, R88 (2008).
110. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
111. Wu, D. et al. A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* **462**, 1056–1060 (2009).
112. Wu, D., Jospin, G. & Eisen, J. A. Systematic identification of gene families for use as 'markers' for phylogenetic and phylogeny-driven ecological studies of Bacteria and Archaea and their major subgroups. *PLoS ONE* **8**, e77033 (2013).
113. Eddy, S. R. Hidden markov models. *Curr. Opin. Struct. Biol.* **6**, 361–365 (1996).
114. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29–W37 (2011).
115. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
116. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
117. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
118. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
119. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
120. Löytynoja, A. & Goldman, N. An algorithm for progressive multiple alignment of sequences with insertions. *Proc. Natl Acad. Sci. USA* **102**, 10557–10562 (2005).
121. Löytynoja, A. & Goldman, N. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* **320**, 1632–1635 (2008).
122. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577 (2007).
123. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
124. Huerta-Cepas, J., Dopazo, J. & Gabaldón, T. ETE: a python environment for tree exploration. *BMC Bioinformatics* **11**, 24 (2010).
125. Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* **33**, 1635–1638 (2016).
126. Cock, P. J. A. et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).
127. Popescu, A. A., Huber, K. T. & Paradis, E. Ape 3.0: new tools for distance-based phylogenetics and evolutionary analysis in R. *Bioinformatics* **28**, 1536–1537 (2012).
128. Revell, L. J. Phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2011).
129. Pateiro-López, B. & Rodrıguez-Casal, A. Generalizing the convex hull of a sample: the R package alphahull. *J. Stat. Softw.* **34**, 1–28 (2010).
130. VanDerWal, J., Falconi, L., Januchowski, S., Shoo, L. & Storlie, C. SDMTools: tools for processing data associated with species distribution modelling exercises. R version 1 (2014); https://cran.r-project.org/web/packages/SDMTools/index.html
131. Adler, D. et al. Rgl: 3D visualization using OpenGL. R version 095 (2016); https://cran.r-project.org/web/packages/rgl/index.html
132. Ross, K. F. A. & Billing, E. The water and solid content of living bacterial spores and vegetative cells as indicated by refractive index measurements. *Microbiology* **16**, 418–425 (1957).
133. Lee, J. A. et al. MIFlowCyt: the minimum information about a flow cytometry experiment. *Cytometry A* **73**, 926–930 (2008).

134. Spidlen, J., Breuer, K., Rosenberg, C., Kotecha, N. & Brinkman, R. R. FlowRepository: a resource of annotated flow cytometry datasets associated with peer-reviewed publications. *Cytometry A* **81**, 727–731 (2012).
135. Barrick, J. E. et al. Identifying structural variation in haploid microbial genomes from short-read resequencing data using *breseq*. *BMC Genom.* **15**, 1039 (2014).
136. Deatherage, D. E. & Barrick, J. E. in *Engineering and Analyzing Multicellular Systems: Methods and Protocols* (eds Sun, L. & Shou, W.) 165–188 (Springer, 2014).
137. Zhu, B. & Stülke, J. *Subti*Wiki in 2018: from genes and proteins to functional network annotation of the model organism *Bacillus subtilis*. *Nucleic Acids Res.* **46**, D743–D748 (2018).

## Acknowledgements

J.v.G. thanks M. Olombrada, M. Toll-Riera, N. Lyons and J. Payne for discussions. We thank R. Kolter for providing strains. J.v.G. thanks the Wierenga Rengerink PhD Prize from the University of Groningen, Rubicon Fellowship (No. 2015-2) from the Netherlands Organisation for Scientific Research (NWO), EMBO long-term fellowship (ALTF, No. 1101-2016), Marie Sklodowska-Curie Individual Fellowship (No. 742235), Swiss Federal Institute of Aquatic Science and Technology (Eawag) and ETH Zürich for financial support. A.W. acknowledges support by ERC Advanced Grant No. 739874, Swiss National Science Foundation grant No. 31003A_172887 as well as by the University Priority Research Program in Evolutionary Biology at the University of Zurich. M.A. was supported by Swiss National Science Foundation grant No. 31003A_169978, Eawag and ETH Zürich.

## Author contributions

J.v.G. conceived the project. All authors were involved in designing the project, planning the research and interpreting the results. J.v.G. conducted the research, created the figures and wrote a first version of the manuscript. M.A. and A.W. contributed to subsequent versions of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at https://doi.org/10.1038/s41559-019-0939-6.



**Correspondence and requests for materials** should be addressed to J.v. or A.W.

nature research

Corresponding author(s): Jordi van Gestel

Last updated by author(s): May 20, 2019

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☒ | ☐ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | Data was analyzed using R version 3.4.3, Matlab R2016a, Python 3.4.4 |
| Data analysis | All software packages and libraries that were used for data analysis are specified in the Methods. |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The flow cytometry data is publically available through the FlowRepository, accession number: FR-FCM-ZYVN (http://flowrepository.org/id/FR-FCM-ZYVN). The sequencing reads are publically available on the European Nucleotide Archive (ENA) database, accession number: PRJEB32792 (http://www.ebi.ac.uk/ena/data/view/PRJEB32792). The remaining data are included in the Supplementary Information or available through public repositories as mentioned in the Supplementary Information.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | Sample size was determined based on the largest number of bacterial populations that could be handled during the evolution experiment and subsequent phenotyping. |
| Data exclusions | No data was excluded from this study. |
| Replication | All phenotypes are characterized using common garden experiments in multiple independent replicates (the exact number of replicates is specified in the figure captions). All replicates gave qualitatively similar results. |
| Randomization | Randomization was not relevant for this study, since all bacterial populations were exposed to the same experimental treatment. |
| Blinding | Blinding was not possible in the phenotypic analysis, because strains and species can be recognized in phenotyping (incl. microscopy, flow cytometry and colony growth). |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ Antibodies |
| ☒ | ☐ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology |
| ☒ | ☐ Animals and other organisms |
| ☒ | ☐ Human research participants |
| ☒ | ☐ Clinical data |

## Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☐ | ☒ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |

# Flow Cytometry

## Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

## Methodology

| | |
|---|---|
| Sample preparation | Cells were collected from the edge or center of the colonies using a 10uL inoculation loop and resuspended in PBS buffer. These cell suspensions were subsequently diluted to an OD600 of 0.001. The normalized cell suspension was stained using SYBR Green I staining according to the protocol described in the methods and subsequently characterized using the flow cytometer. |
| Instrument | BD Accuri C6 flow cytometer |
| Software | Flow cytometer data was analyzed in R version 3.4.3 |
| Cell population abundance | Cell abundances after gating are shown in Fig. 6d and Supplementary Fig. 31 |

Gating strategy | The gating strategy is specified in the Methods and illustrated in Fig. 6b and Supplementary Fig. 28

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.